

Entwicklung informatischer, mathematischer und instrumenteller Methoden zur Analyse physiologischer Prozesse in mikrobiellen Populationen

Isam Haddad



**Technische
Universität
Braunschweig**



**Entwicklung informatischer,
mathematischer und instrumenteller
Methoden zur Analyse
physiologischer Prozesse in
mikrobiellen Populationen**

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigte
D i s s e r t a t i o n

von Isam Haddad
aus Damaskus, Syrien

1. Referent: Professor Dr. Dieter Jahn
2. Referent: Privatdozent Dr.-Ing. Max Schobert
eingereicht am: 18.06.2012
mündliche Prüfung (Disputation) am: 31.08.2012

Druckjahr 2012

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Haddad, I., Münch, R., Jahn, D. (2010). Vorrichtung und System zur Durchführung von zeitaufgelösten Untersuchungen an Zellproben. Patentanmeldung am Deutschen Patent- und Markenamt (DE: 10 2010 048 443.1)

Tielen, P., Narten, M., Rosin, N., Biegler, I., **Haddad, I.**, Hogardt, M., Neubauer, R., Schobert, M., Wiehlmann, L. & Jahn, D. (2010). Genotypic and phenotypic characterization of *Pseudomonas aeruginosa* isolates from urinary tract infections. Int. J. Med. Microbiol., 301: 282-292

Haddad, I., Hiller, K., Frimmersdorf, E., Benkert, B., Schomburg, D., Jahn, D. (2009). An emergent self-organizing map based analysis pipeline for comparative metabolome studies. In silico biology. 9(4):163-78.

Grote, A., Klein, J., Retter, I., **Haddad, I.**, Behling, S., Bunk, B., Biegler, I., Yarmolinetz, S., Jahn, D. & Münch, R. (2009). PRODO-RIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes. Nucleic Acids Res. Database issue D61-5

Klein, J., Münch, R., Biegler, I., **Haddad, I.**, Retter, I. & Jahn, D. (2009). Strepto-DB, a database for comparative genomics of group A (GAS) and B (GBS) *Streptococci*, implemented with the novel database platform 'Open Genome Resource' (OGeR) Nucleic Acids Res. Database issue D494-8

Pommerenke, C., Gabriel, I., Bunk, B., Münch, R., **Haddad, I.**, Tielen, P., Wagner-Döbler, I., Jahn, D. (2008). ROSY – a flexible and universal database and bioinformatics tool platform for *Roseobacter* related species. In silico biology. 8(2):177-86

Tagungsbeiträge

Haddad, I., Münch, R., Jahn, D. (2010). Dynamic bacteriochlorophyll measurement in *Dinoroseobacter shibae*. Gastvortrag, Symposium des DFG Sonderforschungsbereiches/Transregio 51, Delmenhorst

Haddad, I., Hiller, K., Frimmersdorf, E., Benkert, B., Schobert, M., Schomburg, D., Jahn, D. (2008). eSOMet: A novel tool for highly reliable biomarker identification Poster, International Conference on Systems Biology, Göteborg, Schweden.

Klein, J., Retter, I., Leupold, S., Grote, A., **Haddad, I.**, Jahn, D., Münch, R. (2008). PRODORIC - a platform for the elucidation of gene regulatory networks in prokaryotes. Poster, International Conference on Systems Biology, Göteborg, Schweden.

Yarmolinets, S., **Haddad, I.**, Retter, I., Klein, J., Jahn, D., Münch, R. (2011). Analysis of the structure and evolutionary conservation of regulatory networks in bacteria. Poster, International Conference on Intelligent Systems for Molecular Biology and European Conference on Computational Biology, Wien, Österreich

Inhaltsverzeichnis

Vorveröffentlichungen der Dissertation	5
Abkürzungen und Akronyme	13
1 Einleitung	17
1.1 Modellbildung in den Lebenswissenschaften	17
1.1.1 Bioinformatik und Systembiologie	17
1.1.2 Hochdurchsatztechnologien	20
1.1.3 Modellierungsansätze	22
1.2 Metabolomanalysen	25
1.2.1 Instrumentelle Analytik	25
1.2.2 Multivariate Statistik	27
1.2.3 Methoden künstlicher neuronaler Netze	30
1.2.4 Relevante Datenbanken	33
1.3 <i>Dinoroseobacter shibae</i>	33
1.3.1 Allgemeine Charakterisierung	33
1.3.2 Aerobe anoxygenen Photosynthese	35
1.4 Evolutionäre Spieltheorie	36
1.4.1 Heterogenität in isogenetischen Populationen	36
1.4.2 Populationsdynamik	38
1.4.3 Replikatordynamik	38
1.5 Zielsetzung	43

2	Materialien und Methoden	45
2.1	Methoden der Software-Entwicklung	45
2.1.1	Die Entwicklungswerkzeuge	45
2.1.1.1	Programmierung in Java	45
2.1.1.2	Programmierung in C#	45
2.1.1.3	Programmierung und Datenauswertung in R	46
2.1.1.4	Verwendete Bibliotheken	46
2.1.2	Die KEGG Datenbank	46
2.1.3	Biostatistische Methoden	47
2.1.3.1	Multivariate Datenmatrix	47
2.1.3.2	Datennormalisierung	48
2.1.3.3	Hauptkomponentenanalyse	49
2.1.3.4	Hierarchische Clusteranalyse	50
2.1.3.5	Selbstorganisierende Karten	51
2.1.3.6	Dendrogramme	54
2.1.3.7	Silhouettenbreite	54
2.1.3.8	Tests der schließenden Statistik	55
2.2	Experimentelle Methoden	56
2.2.1	Geräte	56
2.2.2	Chemikalien und Verbrauchsmaterialien	57
2.2.3	Kultivierungsbedingungen von <i>D. shibae</i>	57
2.2.3.1	Bakterienstamm	57
2.2.3.2	Medien	57
2.2.3.3	Anzucht aerober Vorkulturen	60
2.2.3.4	Kultivierung in Mikrotiterplatten	60
2.2.3.5	Lichtbedingungen während der Mikrotiterplattenkultivierung	62
2.2.3.6	Synchrone Steuerung der Mikrotiterplattenbeleuchtung	63
2.2.4	Bacteriochlorophyll Visualisierung und Quantifizierung	64
2.2.4.1	Licht- und Fluoreszenzmikroskopische Untersuchungen	64

2.2.4.2	<i>In vivo</i> -Quantifizierung von Bacteriochlorophyll . . .	66
3	Ergebnisse und Diskussion	69
3.1	eSOMet – Eine neue Software zur Metabolomdatenanalyse	69
3.1.1	Ein verbessertes Verfahren zur Metabolomdatenauswertung	69
3.1.2	Das Softwaredesign	72
3.1.3	Evaluierung von eSOMet mit einem Metabolomdatensatz von <i>Corynebacterium glutamicum</i> Kultivierungen	76
3.1.4	Quantitativer Vergleich der verschiedenen Clusterverfahren	83
3.2	Lichtabhängige Physiologie und Morphologie von <i>D. shibae</i>	86
3.2.1	Populationsdynamik unter fluktuierenden Lichtbedingungen	86
3.2.1.1	Motivation eines spieltheoretischen Modellierungsansatzes	86
3.2.1.2	Definition von Zustands- und Ereignisraum	89
3.2.1.3	Rücktransitionen im Stufen-Matrix Modell durch diskretes Wachstum	92
3.2.1.4	Übergang des Stufen-Matrix Modells in kontinuierliches Wachstum	97
3.2.1.5	Stochastische Zufallsverteilung	99
3.2.1.6	Vorwärtstransitionen im Stufen-Matrix Modell durch Produktion	100
3.2.1.7	Einbettung in die Replikatorgleichung	102
3.2.1.8	Integration von Umweltreizen und Simulation der <i>Roseobacter</i> Hypothese	103
3.2.1.9	Experimentelle Modellverifikation	107

3.2.2	Mikroreaktionstechnische Kultivierung von <i>D. shibae</i>	109
3.2.2.1	Vorüberlegungen und Technik	109
3.2.2.2	Etablierung der relativen BChl- <i>a</i> Quantifizierung . . .	110
3.2.2.3	Verdunstungseffekte und Schichtdickenkorrektur . . .	115
3.2.2.4	Etablierende Wachstumsexperimente	117
3.2.3	Hell-Dunkel-Bedingungen im Mikroreaktionsmaßstab	121
3.2.3.1	Die Entwicklung eines Prototypen	121
3.2.3.2	<i>In-vivo</i> BChl- <i>a</i> Messungen unter Hell-Dunkel-Bedingungen	125
3.2.3.3	Konzeptionierung eines fortgeschrittenen Prototypen .	127
4	Zusammenfassung und Ausblick	131
4.1	Zusammenfassung	131
4.2	Ausblick	133
	Literaturverzeichnis	135
	Danksagung	155
	Anhang	159
A1	Dendrogramm der ESOM	159
A2	Dendrogramm der HCA	162
A3	Beweis zu Satz 1	165
A3	Beweis zu Satz 2	167

Abkürzungen und Akronyme

bspw.	—	beispielsweise
AAAnP	—	Aerobe Anoxygenic Photosynthese
AI	—	<i>engl.</i> : Artificial Intelligence
ATP	—	Adenosintriphosphat
BChl- <i>a</i>	—	Bacteriochlorophyll- <i>a</i>
BRENDA	—	Braunschweig Enzyme Database
<i>Corynebacterium glutamicum</i>	—	<i>C. glutamicum</i>
CSV	—	<i>engl.</i> : Character Separated Values
deion.	—	deionisiert
DIKW	—	<i>engl.</i> : Data-Information-Knowledge-Wisdom
<i>D. shibae</i>	—	<i>Dinoroseobacter shibae</i>
ESOM	—	<i>engl.</i> : Emergent Self Organizing Map
ES	—	Evolutionäre Spieltheorie
ESS	—	<i>engl.</i> : Evolutionary Stable State
FTP	—	<i>engl.</i> : File Transfer Protocol
GC-MS	—	Gaschromatographie-Massenspektrometrie
ggf.	—	gegebenenfalls
HCA	—	Hierarchische Cluster Analyse
HID	—	<i>engl.</i> : Human Interface Device
JDK	—	<i>engl.</i> : Java Development Kit
KEGG	—	Kyoto Encyclopedia of Genes and Genomes
KNN	—	Künstliches neuronales Netz

LED	—	<i>engl.</i> : Light-emmiting Diode
LH2	—	<i>engl.</i> : Light-Harvesting Complex 2
MB	—	Marine Boullion
MCA	—	<i>engl.</i> : Metabolic Control Analysis
min	—	Minuten
NIR	—	Nahinfrarot
PCA	—	<i>engl.</i> : Principal Component Analysis
rel.	—	relativ/e/er
rpm	—	<i>engl.</i> : Revolutions Per Minute
rRNA	—	<i>engl.</i> : Ribosomal Ribonucleic Acid
sog.	—	sogenannte/n
SOM	—	<i>engl.</i> : Self Organizing Map
SQL	—	<i>engl.</i> : Structured Query Language
SWM	—	Salzwasser-Medium
URL	—	<i>engl.</i> : Uniform Resource Locator
USB	—	<i>engl.</i> : Universal Serial Bus
usw.	—	und so weiter
u.U.	—	Unter Umständen
UV	—	Ultraviolett
XML	—	eXtensible Markup Language
z.B.	—	zum Beispiel

1 Einleitung

*... while ideas ultimately can be so powerful,
they begin as fragile, barely formed thoughts.
So easily missed, so easily compromised, so easily just
squished.*

Jonathan Ive, 2011

1.1 Modellbildung in den Lebenswissenschaften

1.1.1 Bioinformatik und Systembiologie

Die modernen Biowissenschaften vollziehen derzeit einen Paradigmenwechsel. Dieser wurde insbesondere durch das zunehmende Aufkommen neuer *integrativer* Analysemethoden vermittelt, die im Rahmen der klassischen Molekularbiologie die Aufmerksamkeit von reduktionistischen Ansätzen hin zur Betrachtung ganzheitlicher Systeme lenkte (Kitano, 2002). Gleichwohl wäre es nicht richtig anzunehmen, dass integrative Analysen – also solche, die verschiedene Aspekte molekularer Systeme und ihre Wechselwirkung berücksichtigen – eine Neuerung der letzten Jahre wären. Tatsächlich ist die neue Disziplin der Systembiologie nur ein vorläufiges Ende einer lang anhaltenden Evolution zweier beinahe unabhängiger Entwicklungsstränge (Westerhoff & Palsson, 2004).

Der eine Strang lässt sich insbesondere bis zur Entdeckung der DNA-Struktur und ihrer kodierenden Eigenschaften zurückverfolgen (Watson & Crick, 1953). Darauf folgten Erkenntnisse über die rekombinanten Mechanismen der Zellen und deren technischer Adaption. Neue genetische Methoden wie die Klonierung (Arber & Linn, 1969; Danna & Nathans, 1971; Sambrook & Russell, 2001) und die Polymerasekettenreaktion (PCR, Mullis *et al.*, 1986) machten ein gezieltes molekularbiologisches Arbeiten möglich und fanden einen vorläufigen Höhepunkt in der Sequenzierung vollständiger Genome, mit Hilfe derer ein ganzheitlichen Blick auf das physiologische Potential ganzer Organismen freigegeben wurde (Fleischmann *et al.*, 1995; Venter *et al.*, 2001). Mit der Pyrosequenzierung steht heutzutage eine Methode zur Verfügung, mit der sich bakterielle Genome in weniger als 24 h sequenzieren lassen (Chen *et al.*, 2010; Novais & Thorstenson, 2011). Dieser Entwicklungsstrang bereitete die strukturelle Grundlage der Systembiologie.

Parallel dazu entwickelte sich die Forderung, die Fülle der deskriptiven Daten in quantitative, molekularbiologische Zusammenhänge zu setzen. Erste mathematische Modelle beschäftigten sich mit der Frage, wie lebende Systeme den hohen Grad an Ordnung aufrecht erhalten können und damit augenscheinlich die thermodynamische Forderung nach Entropie-Maximierung verletzen (Schrödinger, 1943). Hierauf versuchte das vereinfachte Prinzip der Reziprozitätsbeziehung eine Antwort zu geben (Onsager, 1931), welches die Kopplung verschiedener Flüsse und Kräfte eines thermodynamischen Systems beschreibt, dessen Zustand sich in der Nähe des Gleichgewichts befindet. Als sich im Laufe der Zeit das Potential biologischer Systeme in Form stetig wachsender metabolischer Netzwerke ausdrückte, begannen Bestrebungen thermodynamische Modelle auf diese anzuwenden – ein Vorhaben, welches sicherlich

erst durch die aufkommende Verfügbarkeit von Rechnersystemen in Reichweite geriet (Garfinkel *et al.*, 1970). Das daraus resultierende Konzept der metabolischen Kontrollanalyse (MCA, Heinrich *et al.*, 1977) war einer der frühen Versuche intrazelluläre Prozesse ganzheitlich, integrativ und quantitativ zu beschreiben.

Die Ideen im Kern der Systembiologie reichen soweit zurück, dass es nicht richtig wäre diese Disziplin zu einer sehr jungen oder gar revolutionären Wissenschaft zu erheben, da sie auf dem holistischen Grundgedanken der Systemtheorie fußt. Faktisch konnten aber erst durch die rasante Verbreitung von Computersystemen und die stete Verbesserung analytischer Technologien, die hier beschriebenen Entwicklungsstränge vollständig koevolvieren, was endgültig dazu führte die Quantifizierbarkeit biologischer Prozesse einzufordern (Westerhoff & Palsson, 2004).

Die stetig wachsende Fülle an produzierten Daten molekularbiologischen Ursprungs macht die eigentliche Aufgabe der Systembiologie deutlich: Erst durch das Durchschreiten verschiedener Stufen des Verstehens von Zusammenhängen, Mustern und Prinzipien lassen sich reine (Mess- oder Roh-) Daten über Information und Wissen in Erkenntnis verwandeln. Jeder Übergang erfordert hierbei einen höheren Grad der Einbeziehung der verschiedenen Aspekte des beobachteten Systems (Abbildung 1.1). Dieses wissenschaftstheoretische Prinzip fand unter dem Schlagwort *Data-Information-Knowledge-Wisdom* (DIKW) breite Beachtung und die Motivation dem als Herausforderung gerecht zu werden, ist die treibende Kraft hinter der Systembiologie und deren Teilbereiche der Bioinformatik und Biomathematik.

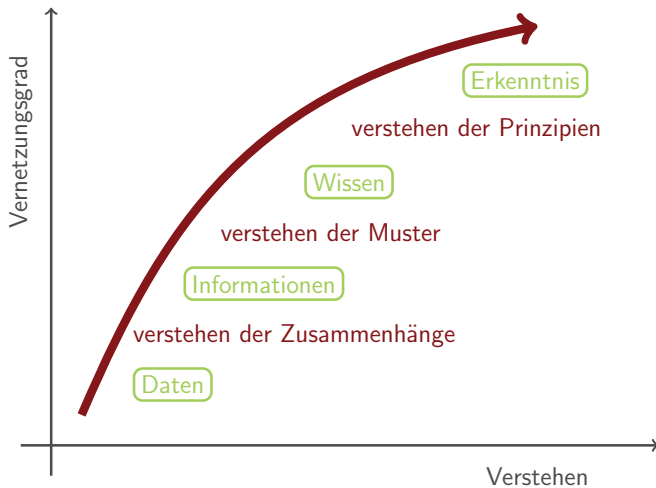


Abbildung 1.1: Vereinfachte Darstellung des DIKW-Prinzips. Schematisch dargestellt sind die Transitionen von Daten über Informationen und Wissen hin zu Erkenntnis. Die Transition von einer Kategorie menschlichen Erfassens zur Nächsten erfordert ein partikuläres Verstehen bestimmter Zusammenhänge und wird durch einen höheren Grad an Vernetzung der bereits erfassten Aspekte vermittelt. (Erstellt nach Bellinger *et al.*, 2004)

1.1.2 Hochdurchsatztechnologien

Wie eingangs beschrieben, war das Aufkommen der Hochdurchsatztechnologien, das was die Systembiologie auf eine substanzielle Basis stellte. Allgemein versteht man darunter eine analytische (meist instrumentelle) Methode, die in der Lage ist den Zustand eines molekularbiologischen *Aspekts* des beobachteten Systems möglichst vollständig zu erfassen. Mit Aspekt wiederum ist die Gesamtheit einer molekularen Organisationseinheit, wie der des Genoms, Transkriptoms, Proteoms oder Metaboloms gemeint. Die quantitative Analyse dieser Aspekte wird durch das Suffix „-omik“ ausgedrückt, also der

Genomik (Analyse der Proteinogenen, regulatorischen und nichtkodierenden Sequenzen), Transkriptomik (Analyse der Genexpression und RNA-Produktion), Proteomik (Analyse der Proteinkonzentrationen) und Metabolomik (Analyse der metabolischen Zusammensetzung). Bilden diese vier Technologien den Kern molekularbiologischer Hochdurchsatzmethoden, entwickelten sich in der Vergangenheit, beinahe inflationär, eine Vielzahl von Technologien bezüglich der peripheren Aspekte, wie z.B. die Phylogenomik (vergleichende Analyse genetischer Aspekte über phylogenetische Zweige), die Physionomik (Analyse physiologischer Dynamiken ganzer Organismen), die Lipomik (Analyse des Fettsäurestoffwechsels), usw.

Alle aufgezählten Technologien variieren in ihren Blickrichtungen auf das beobachtete System. Während insbesondere die der Genomik zugehörigen Analysen Untersuchungen genotypischer Eigenschaften anstreben, sind andere Analysen auf phänotypische Eigenschaften wie z.B. die Reaktion auf Umweltreize ausgelegt. Obwohl also die Fragestellungen immer verschiedene sind und auch die instrumentellen Methoden sich meist sehr stark unterscheiden, folgt der reinen Datenerhebung immer ein bioinformatisches Programm, das aus der Sequenz vierer Teilschritte zusammengesetzt ist (Schneider & Orchard, 2011):

1. Rohdatenprozessierung und Identifikation der Entitäten (als Entität ist hier bspw. das quantifizierte Gen, Transkript, Protein oder Metabolite gemeint).
2. Statistische Auswertung der gefundenen Informationen.
3. Wissensbasierte Abbildung der Informationen auf Netzwerken, um ihre Zusammenhänge zu verstehen.
4. Erkenntnisgewinn aus gewonnenem Wissen durch Modellbildung.

In diesem allgemeingültigen Programm zeichnet sich deutlich das eingangs beschriebene DIKW-Prinzip ab.

1.1.3 Modellierungsansätze

Ein Modell entwickelt sich immer aus dem Versuch eine partikuläre Fragestellung zu einem Phänomen zu beantworten und sollte somit nie einem Selbstzweck folgen. Bereits die Fragestellung präzise zu formulieren und daraus die Wahl eines geeigneten Modellierungsansatzes abzuleiten, sind demnach essentielle Teilschritte des Modellierungsprozesses. Grundsätzlich kann man Modelle in solche, die einen vorhersagenden Charakter und jene, die einen erklärenden Charakter haben unterscheiden. Dabei sollte man berücksichtigen, dass jedes Modell in aller Regel beide Eigenschaften in verschiedenen Gewichtungen vereint: Das Erklären einer Beobachtung lässt in Grenzen stets Vorhersagen zu, wohingegen die Verifikation bzw. Falsifikation einer Vorhersage auch immer eine erklärende Wirkung hat. Gemäß Kell & Knowles, 2006 ergeben sich daraus verschiedene Motivationen für die Modellierung:

- Die Überprüfung, ob ein Modell experimentelle Daten bestätigt.
- Die Analyse des Modells bezüglich der Parameter, die einen nicht zu vernachlässigenden Einfluss auf die Fragestellung haben.
- Hypothesen-Bildung und -Tests am Modell, anstelle des untersuchten Objekts.
- Untersuchungen, welche Veränderungen eines bestehenden Modells die Realität besser abbilden.

Wenn im Rahmen dieser Arbeit von einem Modell gesprochen wird, ist damit immer ein mathematisches Modell gemeint. Insofern ist es sinnvoll einige Begriffsbestimmungen durchzuführen: Gegenstand des Modells ist das (biologische) *System*, welches es abzubilden versucht. Dieses ist durch die Zusammensetzung seiner *Objekte* definiert, wobei ein Objekt sowohl etwas Materielles, als auch eine Eigenschaft sein kann. Die Objekte können verschiedene *Zustände* aufweisen. Die Menge aller Zustände, die das Modell zulässt, wird als *Zustandsraum* bezeichnet. Das Modell bildet die Änderungen der Zustände innerhalb eines geeigneten *Ereignisraums* ab. Der Ereignisraum wird meistens in, von einander unabhängigen, physikalischen Größen wie Zeit oder Raum ausgedrückt.

Weiterhin lässt sich eine Klassifizierung von Modellen nach Kell & Knowles, 2006 durch diejenigen Eigenschaften vornehmen, durch die die oben abstrakt beschriebenen Begriffe eine Realisierung erfahren:

Stochastisch oder Deterministisch Stochastische Modelle berücksichtigen, dass sich die Zustände der Objekte immer mit einer gewissen Wahrscheinlichkeit bzw. über eine Wahrscheinlichkeitsverteilung ändern. Modelle, die als Vereinfachung darauf verzichten, werden deterministisch genannt. Diese Begriffsfindung kann irreführend sein, da sie den Schluss zulässt, dass stochastische Modelle unter gleichen Bedingungen nicht zwingend zum gleichen Ergebnis führen. Zwar ist das für bestimmte Verfahren (z.B. Monte-Carlo-Simulationen oder auf artifizieller Intelligenz basierenden) auch tatsächlich der Fall, gilt aber nicht immer. Insbesondere solche Modelle, die durch stochastische Prozesse ausgedrückt werden, führen durchaus zu deterministischen Ergebnissen, drücken aber das Eintreten der modellierten Zustände durch Wahrscheinlichkeitsverteilungen aus. Deterministische Modelle hingegen ordnen einem Objekt über den Ereignisraum

konkrete, nicht-probabilistische Zustände zu.

Diskret oder Kontinuierlich Kontinuierliche Modelle beschreiben den Zustandsraum immer über den gesamten Ereignisraum, wohingegen der Ereignisraum in diskreten Modellen nur an festgelegten Stellen definiert ist.

Makroskopisch oder Mikroskopische Makroskopische Modelle betrachten ein Objekt immer als Gesamtheit seiner Entitäten. Der Zustand ist dementsprechend immer der Mittelwert der Zustände dieser Entitäten. In makroskopischen Populationsmodellen ist die Population das untersuchte Objekt. Der modellierte Zustand ist häufig die Reproduktionsrate. In mikroskopischen Modellen werden hingegen die Individuen der Population auf Objekten abgebildet, welche bestimmte Zustände annehmen können.

Hierarchisch oder Mehrstufig Hierarchische Modelle zerlegen das beobachtete System in mehrere Teilsysteme mit einem jeweils reduzierten Satz an Objekten und modellieren auf der nächsthöheren Hierarchieebene die Teilsysteme als voneinander abhängige Objekte. Mehrstufigen Modellen hingegen liegt die Annahme zugrunde, dass Interaktionen der Objekte, die in Hierarchien gekapselt wären, nicht vernachlässigt werden dürfen.

Die Auswahl der hier gegenübergestellten Eigenschaften drückt implizit die wichtigste Frage aus, die der Modellierung vorangeht: Wie stark vereinfacht darf ein Modell sein, um das zu untersuchende Phänomen noch exakt genug zu beschreiben? Umgekehrt geht jeder Verzicht von Reduktionismus mit einer wachsenden Komplexität einher, die das Modell ebenso unbrauchbar machen kann. Die Verfügbarkeit bestimmter Daten lässt für jede dieser Eigenschaften

eine Entscheidung zu, die letztendlich zu der Wahl eines geeigneten Modellierungsansatzes führt.

1.2 Metabolomanalysen

1.2.1 Instrumentelle Analytik

Die Metabolomik erfuhr in der vergangenen Dekade eine hohe Aufmerksamkeit, da sie einen unmittelbaren Zugang zu den phänotypischen Eigenschaften von pro- und eukaryotischen Organismen liefert. Anders als die Transkriptomik oder Proteomik, welche auf die quantitative Analyse der Endprodukte regulatorischer Prozesse fokussiert sind, lässt die Metabolomik somit die Untersuchung der finalen, biochemischen Antwort auf Umweltreize zu (Fiehn, 2002; Kopka *et al.*, 2004).

Unter dem *Metabolom* versteht man die Gesamtheit der niedermolekularen Substanzen, die zu einem Zeitpunkt in einer biologischen Probe zu finden sind. Eine Möglichkeit diese simultan und qualitativ zu identifizieren und quantitativ zu bestimmen, wird durch Gas- oder Flüssigchromatographen, die an geeignete Massendetektoren gekoppelt sind, vermittelt (Fiehn & Kind, 2006). Hierzu werden die aus der Probe extrahierten niedermolekularen Substanzen in einem ersten Schritt nach ihren physikalischen Eigenschaften chromatographisch aufgetrennt. Je nach verwendeter Technik ergibt sich für jede Substanz eine verschieden lange Retentionszeit. Diese ist zwar, bezogen auf das verwandte Analysesystem für jede Substanz charakteristisch, für eine eindeutige Identifikation jedoch nicht hinreichend. Daher werden die eluierenden Substanzen kontinuierlich in einen Massendetektoren geleitet. In diesem werden die Abundanzen und je nach Typ des Detektors die molekularen Massen, bzw. bei

fragmentierenden Detektoren die Massenspektren, jeder Substanz ermittelt.

Die so gewonnenen Rohdaten durchlaufen anschließend eine chemo-informatische Auswertung, die die Daten in erster Linie um instrumentell bedingte Messfehler wie, chromatographische Verschiebungen und Rauschen bereinigt, Spektren koeluerender Substanzen dekonvolutiert und letztendlich die Annotation der gefundenen Substanzen halb- oder vollautomatisch unterstützt. Hierzu wird die Annahme getroffen, dass die Retentionszeit gemeinsam mit den Signalen des Massendetektors eindeutig qualifizierend ist.

Im Ergebnis wird so zu jeder Probe das metabolische Profil, also eine Liste von Metaboliten und zugehörigen Abundanzen ermittelt. Dieses stellt daher eine Momentaufnahme der chemischen Zusammensetzung der untersuchten Probe dar und lässt alleine keine bedeutenden Rückschlüsse auf die zu untersuchenden Stoffwechselprozesse zu. Erst der Vergleich zwischen Proben, die unter faktoriellen Änderungen der Umwelt, des Geno- oder Phänotyps oder als Zeitreihe untersucht werden, liefern tiefere Einblicke. Um letztendlich metabolische Flüsse innerhalb verzweigter Netzwerke aufzuklären, werden die Proben in aller Regel mit isotoopenmarkierten Substraten (typischerweise ^{13}C -markiert) inkubiert, um die sukzessiven Stoffumwandlungen nachvollziehen zu können.

Die Abundanzen der metabolischen Profile sind in aller Regel nicht *voll-quantitativ*. Sie stellen also ohne eine Kalibrierung keine dimensionsbehafteten Konzentrationen dar. Allerdings sind die gewonnen Messwerte *semi-quantitativ*; also die dimensionslosen Abundanzen direkt proportional zur tatsächlichen Konzentration. Damit lassen sich relative Änderungen der Metabolite über die metabolischen Profile verschiedener Proben zuverlässig bestimmen. Für typische Metabolomexperimente ist dies in aller Regel voll ausreichend.

1.2.2 Multivariate Statistik

Einem Metabolomexperiment geht typischerweise die Fragestellungen voraus, wie sich der Stoffwechsel bzw. die metabolische Zusammensetzung des biologischen Systems unter verschiedenen Bedingungen verändert. Die Änderungen lassen dann wiederum Rückschlüsse auf die physiologischen und regulatorischen Prozesse des Systems zu. Oftmals besteht ein weitergehendes Interesse an der Identifikation solcher Metabolite (sog. *Biomarkern*), deren Abundanz ein signifikanter Indikator für einen bestimmten Phänotyp oder bestimmte Umweltbedingungen ist. Zur Beantwortung solcher Fragen wurde in der Vergangenheit ein breites Spektrum statistischer Methoden entwickelt.

Ausgangspunkt dieser Methoden ist immer die *multivariate Datenmatrix*, die sich aus den metabolischen Profilen generieren lässt. Jede Zeile dieser Matrix repräsentiert einen Metaboliten (statistische Variable); die Spaltenvektoren das metabolische Profil (statistisches Objekt). Stellt man sich vor, dass die Metabolite einen hochdimensionalen Raum aufspannen, lässt sich jeder Spaltenvektor als ein Punkt innerhalb dieses Raumes abbilden (Abbildung 1.2).

Gegenstand der statistischen Auswertung ist es, die Variabilität innerhalb dieser Punktwolke zu erklären. Diese setzt sich letztendlich aus drei Arten zusammen (van den Berg *et al.*, 2006):

Induzierte biologische Variabilität wird durch verschiedene Umwelt- oder genetische Faktoren, die der Probe zugrunde lagen, erklärt.

Nichtinduzierte biologische Variabilität wird durch natürliche, stochastische Fluktuationen im Stoffwechsel erklärt.

Statistisches Rauschen wird durch zufällige und systematische Fehler der experimentellen und technischen Handhabung der Proben erklärt.

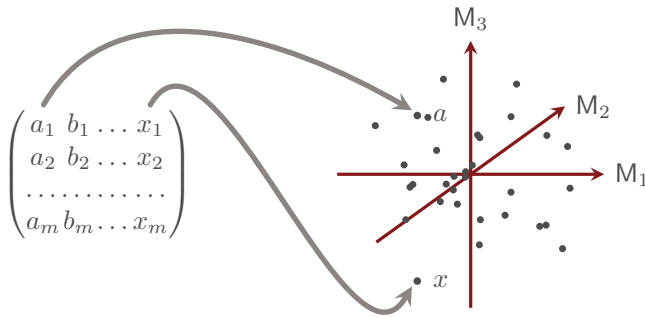


Abbildung 1.2: Mathematische Abstraktion der multivariaten Datenmatrix aus einem Metabolomexperiment. Links dargestellt ist eine beliebige Matrix. Der Eintrag a_2 entspricht der gemessenen Abundanz des Metaboliten M_2 in der Probe a . Die Menge aller statistischen Objekte (Spalten der Matrix) lassen sich so (rechts) auf einem hochdimensionalen Raum abbilden, der durch die statistischen Variablen (hier also die drei Metabolite M_1 , M_2 und M_3) aufgespannt wird.

Grundsätzlich erwartet man, dass der Einfluss der induzierten Variabilität auf die Änderungen der metabolischen Profile größer ist als der, der nichtinduzierten, während das statistische Rauschen den geringsten Einfluss hat. Sowohl die Komplexität und hohe Dimensionalität der Daten, als auch das Auftreten unerwünschter Variabilitätsquellen zeigen die Notwendigkeit auf, der ursprünglichen Fragestellung mit geeigneten mathematischen Methoden zu begegnen. Hierzu spielten in der Vergangenheit insbesondere die folgenden Methoden eine wegweisende Rolle.

Hauptkomponentenanalyse Die Hauptkomponentenanalyse (PCA, *engl.*: Principle Component Analysis) transformiert die Daten durch lineare, algebraische Operationen aus ihrem hochdimensionalen Raum

auf einen niedriger Dimensionalität. Dazu werden diejenigen orthogonalen Basisvektoren berechnet, die entlang der größten Varianz innerhalb der Daten orientiert sind. Derjenige Basisvektor entlang der größten Varianz wird als *erste Hauptkomponente*, der entlang der zweitgrößten Varianz als *zweite Hauptkomponente*, usw. bezeichnet. Eine Projektion der Daten auf die ersten Hauptkomponenten hat somit nicht nur eine komplexitäts- bzw. dimensionsreduzierende Wirkung, sondern entfernt gleichzeitig das uncharakteristische Rauschen nichtinduzierter Variabilität. Gleichzeitig erlaubt die Prozedur die Identifikation der Metabolite, die den größten Beitrag zur Orientierung der Hauptkomponenten und somit den höchsten Grad an Variabilität hatten. Die PCA wurde aus diesen Gründen in der Vergangenheit immer wieder erfolgreich im Rahmen von Metabolomdatenauswertungen eingesetzt (Roessner *et al.*, 2001; Fiehn, 2001; Askenazi *et al.*, 2003; Hirai *et al.*, 2004; Jonsson *et al.*, 2004).

Clusteranalysen Grundsätzlich unterscheidet man zwei Typen von Clusteralgorithmen: Bei den partitionierenden Verfahren wird zu Beginn die zu erwartende Anzahl der Cluster festgelegt und anschließend werden in einem iterativen Prozess die metabolischen Profile den Clustern so zugeordnet, dass die Summe der Distanzen der Profile innerhalb eines Cluster minimiert wird. Häufig eingesetzte Verfahren sind der k-means- oder Expectation-Minimization-Algorithmus (MacQueen, 1967; Dempster *et al.*, 1977; Backhaus *et al.*, 2011). Im Gegensatz dazu verfolgen die agglomerativen Verfahren einen *bottom-up* Ansatz. Es werden zunächst paarweise solche Profile identifiziert, die die höchste Ähnlichkeit haben und zu einem Cluster zusammengefasst. Anschließend werden die Cluster, die untereinander am ähnlichsten sind, sukzessive zu Clustern höherer Ordnung vereinigt bis zuletzt nur ein Cluster übriggeblieben ist. Als Ergeb-

nis erhält man eine hierarchische Ordnung, die sich geeigneterweise als Dendrogramm darstellen lässt (Eisen *et al.*, 1998; Herrero *et al.*, 2001). Beide Verfahrenstypen haben aber auch inhärente Nachteile: Für die Partitionierenden ist die Wahl der ursprünglich festgelegten Anzahl der zu erwartenden Cluster kritisch und kann bei einer fehlerhaften Wahl der Struktur der Daten widersprechen. Im Fall der agglomerativen Verfahren ist der kritische Schritt hingegen die Definition der Ähnlichkeit Cluster höherer Ordnung, die zu teils sehr unterschiedlichen Ergebnissen führen kann.

Varianzanalyse Unter den Varianzanalysen wird eine Sammlung statistischer Methoden zusammengefasst, mit deren Hilfe sich solche Variablen (bezogen auf Metabolomdaten also Metabolite) identifizieren lassen, deren Abundanzen zwischen verschiedenen Einflussfaktoren signifikant variiert sind. Testverfahren der induktiven Statistik, wie der t-Test, liefern hierzu die Aussage mit welcher statistischen Sicherheit (Signifikanz) die Mittelwerte gemessener Abundanzen eines Metaboliten aus Proben verschiedener Teilexperimente voneinander verschieden sind.

1.2.3 Methoden künstlicher neuronaler Netze

Künstliche Neuronale Netze (KNNe) sind eine Abstraktion von Algorithmen, die sich zur Mustererkennung in multivariaten Daten eignen. Motiviert durch Überlegungen der artifiziellen Intelligenz (AI) verarbeiten KNNe Signale ähnlich wie die Nervenzellnetze von Wirbeltiergehirnen. Hierbei wird jedes zu bewertende Signal (z.B. ein metabolisches Profil) über eine Eingabeschicht einer Reihe künstlicher Neuronen „vorgeführt“. Jedes Neuron bewertet das Signal individuell hinsichtlich seiner Eigenschaften und veranlasst für jeden Knoten auf einer Ausgabeschicht eine Bewertung. Im Fall mehr-

schichtiger Netze kann diese Ausgabeschicht wiederum als Eingabe tiefergelegener Schichten fungieren.

Entscheidend für die Signalverarbeitung ist neben der Architektur des Netzes auch die Bewertungsfunktion, die jedem künstlichen Neuron antrainiert werden muss. Hierbei werden maßgeblich zwei Trainingstypen unterschieden (Backhaus *et al.*, 2010):

Überwachte KNNe werden zunächst mit Signalen, deren Bewertung aus anderen Quellen (wie z.B. menschlichen Wissen) bekannt sind, trainiert. Sie eignen sich damit vor allem zur Zuordnung unbekannter Muster zu solchen die bekannt sind.

Unüberwachte KNNe trainieren sich auf Basis der vorliegenden Daten selber. Damit führen sie ähnlich wie Clusteralgorithmen eine Diskriminierung der Signale nach ihrer Ähnlichkeit durch und erzeugen somit ein konsistentes Verarbeitungsmuster.

Selbstorganisierende Karten (SOMs, *engl.*: Self Organizing Maps) sind ein spezieller Typ von einschichtigen KNNe die sowohl überwacht wie auch unüberwacht sein können. Hierbei ist die Ausgabeschicht ein zwei-dimensionales Gitter auf dem jeder Knoten des Gitters mit jedem Neuron verbunden wird. Die dem Netz antrainierte Bewertungsfunktion aktiviert immer dann zwei Knoten in naher Nachbarschaft, wenn die zwei auslösenden Signale eine hohe Ähnlichkeit aufweisen. Zwei Signale werden auf entfernten Knoten aktiviert, wenn deren Muster sich unähnlich sind (Abbildung 1.3). Daher werden SOMs auch als *topologieerhaltend* bezeichnet (Kohonen, 2001).

Ähnlich den partitionierenden Clusterverfahren ist im Fall der SOMs die Größe der Ausgabeschicht von entscheidender Bedeutung: Eine zu klein gewählte Anzahl von Knoten auf dem Gitter kann die innere Struktur der Eingangsdaten falsch wiedergeben. Um diesen Umstand zu beheben, wurde das Konzept der emergenten SOMs

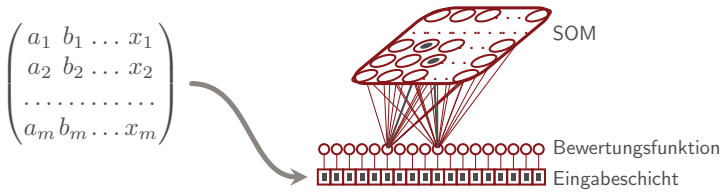


Abbildung 1.3: Schematische Darstellung der Funktionsweise einer SOM. Die Objekte der multivariaten Datenmatrix bilden die Eingabeschicht. Über eine Bewertungsfunktion wird jedes Neuron der Eingabeschicht mit jedem Knoten der trainierten Ausgabeschicht verbunden. Die Bewertungsfunktion vergibt an denjenigen Knoten der Ausgabeschicht die beste Bewertung, dessen antrainiertes Muster dem des Eingangssignals am ähnlichsten ist. Da das Training topologierhaltend ist, werden Eingangssignale mit ähnlichem Muster auf ähnlichen Arealen der SOM abgebildet.

(ESOMs) entwickelt (Ultsch, 1999). Der grundlegende Unterschied ist, dass die Anzahl der Knoten auf der Ausgabeschicht die Anzahl der Eingangssignale um mehrere Größenordnungen übersteigt. Hierdurch wird nach dem abgeschlossenen Training des Netzes den Eingangssignalen der Raum gelassen sich gemäß ihrer Ähnlichkeit auf der Ausgabeschicht zu positionieren. Durch die U*-Matrix Methode (Ultsch, 2003) lässt sich außerdem die gesamte Ausgabeschicht wieder in ein Dendrogramm überführen, welches die globalen Ähnlichkeiten in den ursprünglichen Daten wiedergibt. Gewöhnliche SOMs wurden in der Vergangenheit immer wieder erfolgreich für verschiedene bioinformatische Auswertungen eingesetzt (Tamayo *et al.*, 1999; Kanaya *et al.*, 2001; Abe *et al.*, 2003; Hirai *et al.*, 2004; Meinicke *et al.*, 2008).

1.2.4 Relevante Datenbanken

Mit Hilfe der multivariaten Methoden lassen sich die Daten der metabolischen Profile klassifizieren und differentielle Änderungen in den Abundanzen einzelner Metabolite über die detektierten Klassen identifizieren. Um aber aus den gewonnenen Informationen gemäß dem DIKW-Prinzip tiefergehendes Wissen zu schöpfen, ist ein weitergehendes Verständnis der Interaktionen einzelner Metabolite und der Kontext metabolischer Zusammenhänge nötig. Hierzu kann auf eine Fülle bestehenden Wissens über die biochemischen Prozesse in einer Vielzahl von untersuchten Organismen zurückgegriffen werden. Dieses Wissen ist idealerweise in Datenbanken wie der Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2012), EcoCyc (Keseler *et al.*, 2011), MetaCyc (Caspi *et al.*, 2012) oder der Braunschweig Enzyme Database (BRENDA) (Scheer *et al.*, 2011) hinterlegt. Der Hintergrund dieser Datenbanken bietet eine ideale Grundlage, auf der sich aus den bioinformatischen Auswertungen metabolischer Profile die Muster zugrundeliegender regulatorischer Prozesse ableiten lassen.

1.3 *Dinoroseobacter shibae*

1.3.1 Allgemeine Charakterisierung

Dinoroseobacter shibae ist ein marines, kokkoides oder eiförmiges α -3-Proteobacterium. Seine typischen Abmessungen betragen $0,3 - 0,7 \mu\text{m} \times 0,3 - 1,0 \mu\text{m}$. Das auffälligste Charakteristikum ist die dunkelrot- bis rosafarbene Pigmentierung solcher Zellen, die unter Lichtausschluss gewachsen sind. Die Bakterienart ist dem Roseobacter Cluster innerhalb der Familie der *Rhodobacteraceae* zuzuordnen. Die Vertreter dieses Clusters kommen innerhalb der Weltmeere in ei-

ner Vielzahl von Habitaten vor, wobei ein vermehrtes Auftreten im Polarmeer und küstennahen Regionen dokumentiert ist (Wagner-Döbler & Biebl, 2006; Buchan *et al.*, 2005). Darüber hinaus lassen sich Mitglieder des Clusters regelmäßig in mikrobiellen Gemeinschaften wiederfinden. Insbesondere das Wachstum auf Algen in Form adhärenter Biofilme wurde als erfolgreiche Lebensweise beschrieben (Bruhn *et al.*, 2007).

Der Stamm *D. shibae* DFL12^T wurde von Zellen des Dinoflagellaten *Prorocentrum lima* isoliert (Biebl *et al.*, 2005). Phylogenetische Analysen zeigten, dass der Stamm der Gruppe der *Roseobacter-Sulfitobacter-Silicibacter* zuzuordnen ist und eine hohen Verwandtschaftsgrad zu den Arten *Jannaschia helgolandis*, *Ruegeria atlantica* und *Rhodobacter veldkampii* aufweist. Die Homologie auf Basis von 16S rRNA Sequenzvergleichen betrug hierbei 94,1%, 93,7% und 93,4% (Biebl *et al.*, 2005). Die vollständige Sequenzierung und manuelle Annotation des Genoms konnten in jüngster Vergangenheit abgeschlossen werden und gaben den Blick auf eine Vielzahl putativer Eigenschaften frei (Wagner-Döbler *et al.*, 2010).

Wie alle Mitglieder der Roseobactergruppe ist *D. shibae* heterotroph und weist eine hohe metabolische Vielfalt auf. So kann es ein breites Spektrum organischer Substanzen wie Acetat, Succinat, Fumarat, Laktat, Citrat, Glutamat, Pyruvat, Fruktose und Glycerol als einzige Kohlenstoffquelle zum Wachstum nutzen (Fürch *et al.*, 2009). Im Rahmen der Genomannotation konnten außerdem Gene gefunden werden, die auf die Fähigkeit zum anaeroben Stoffwechsel durch Denitrifikation und Arginin-Fermentation schließen ließen, was sich auch experimentell bestätigen ließ (Piekarski, 2011). Letztendlich zeigt *D. shibae* die Fähigkeit zur aeroben anoxygenen Photosynthese (AAnP), auf die im Folgenden detaillierter eingegangen wird.

1.3.2 Aerobe anoxygenen Photosynthese

In-situ Messungen zeigten eine überraschend hohe und ubiquitäre Abundanz von Bacteriochlorophyll-*a* (BChl-*a*) in der obersten Gewässerschicht der Weltmeere. Dessen Ursprung ist photosynthetisch kompetenten aeroben anoxygen-phototrophen Bakterien zuzuordnen (Kolber *et al.*, 2001). Durch ihre BChl-*a* -basierten Photosysteme sind sie in der Lage, unter aeroben Bedingungen die Energie des Sonnenlichtes in einen Protonengradienten umzuwandeln, der zur Adenosintriphosphat- (ATP) Gewinnung genutzt werden kann (Yurkov & Beatty, 1998). Anders als bei Pflanzen oder Cyanobakterien geschieht dies nicht unter Freisetzung von Sauerstoff. Ihr Anteil an den mikrobiellen Gemeinschaften in entsprechenden Habitaten wird je nach Studie auf 1% - 25% geschätzt (Béjà *et al.*, 2002; Cottrell *et al.*, 2006; Jiao *et al.*, 2010; Oz *et al.*, 2005). Der Beitrag dieser Organismen an dem photosynthetischen Elektronentransport in dieser Gewässerschicht wird auf 2% - 5% geschätzt (Kolber *et al.*, 2000). Durch diese verhältnismäßig hohen Beiträge wird vermutet, dass die AAnP eine wichtige Determinante im globalen Kohlenstoffkreislauf ist (Jiao *et al.*, 2010).

Arten, die die AAnP durchführen entwickelten sich wahrscheinlich aus photosynthetischen Purpurbakterien und stellen eine evolutionäre Zwischenstufe zu nicht-photosynthetisierenden Mikroorganismen dar (Kolber *et al.*, 2000). Während allerdings Purpurbakterien nur unter anaeroben Bedingungen zur autotrophen Photosynthese fähig sind, führen Vertreter der AAnP die Photosynthese ausschließlich unter Anwesenheit von Sauerstoff durch, der für die Funktion des Photosynthese-Apparates benötigt wird (Yurkov & Beatty, 1998; Koblížek *et al.*, 2010). Zudem sind sie photo-heterotroph, verfügen also nicht über die Möglichkeiten Kohlenstoff aus CO₂ zu fixieren. Die durch die Photosynthese gewonnene Energie stellt für die Or-

ganismen einen fakultativen Mehrwert dar, der vermutlich in den oberen, kohlenstoffärmeren aber sauerstoffreichen Gewässerschichten einen Wachstumsvorteil bietet (Buchan *et al.*, 2005).

Bemerkenswerterweise konnte wiederholt gezeigt werden, dass die Produktion von BChl-*a* bei bereits sehr niedrigen Lichtintensitäten gestoppt wird (Yurkov & Gernerden, 1993; Biebl & Wagner-Döbler, 2006; Tomasch *et al.*, 2011). Dies scheint auf den ersten Blick paradox, da die Zellen erst unter hellen Bedingungen von dem produzierten BChl-*a* profitieren können. Durch das Einstellen der Produktion und eine unter Lichtbedingungen erhöhte Teilungsrate kommt es zwangsweise zu einer Verdünnung der intrazellulären BChl-*a* Konzentrationen (Biebl & Wagner-Döbler, 2006). Als Ursache für dieses Verhalten wird vermutet, dass durch die Photosynthese als Nebenprodukt Singulett-Sauerstoff entsteht, der auf die Zellen toxisch wirkt. Das Herunterregulieren der BChl-*a* Produktion kann demnach ein Schutzmechanismus sein, der oxidativem Stress entgegenwirkt (Borland *et al.*, 1989).

1.4 Evolutionäre Spieltheorie

1.4.1 Heterogenität in isogenetischen Populationen

Typische Hochdurchsatztechnologien sind darauf konzentriert, die biologisch induzierte Varianz makroskopisch (durch verschiedene Einflussfaktoren hervorgerufen) aufzuklären (Kapitel 1.2.2). Während stete Verbesserungen der Technologie darauf zielen, die störende instrumentelle Varianz zu minimieren, stellt sich die Frage, welchen Ursprung die nicht-biologisch induzierte Varianz hat. Eingangs wurde bereits erwähnt, dass diese durch stochastische Fluktuationen in den zellulären Prozessen hervorgerufen wird, was aber weder den

Ursprung dieser Fluktuationen, noch ihre mikroskopische Reichweite erklärt.

Bereits frühe Beobachtungen (Delbrück, 1945; Spudich & Koshland, 1976) zeigten, dass isogenetische (also genetisch identische) Populationen von Mikroorganismen, auf mikroskopischer Ebene äußerst heterogene Eigenschaften aufweisen können. Um derartige Phänomene zu erklären, wurde angenommen dass es Effektormoleküle geben muss, deren intrazelluläre Konzentration so niedrig ist, dass ihre Abundanz einer breiten stochastischen Verteilung folgt. Als Konsequenz ist die Häufigkeit einer Reaktion, die der Effektor auslöst, ebenfalls stochastisch verteilt (Gillespie, 1977; Grima & Schnell, 2008). Inzwischen ist bekannt, dass die heterogene Verteilung phänotypischer Ausprägungen in isogenetischen Populationen ein typisches Phänomen ist, welches sich auf weite Bereiche des Metabolismus, der Physiologie und Morphologie erstrecken kann (Brehm-Stecher & Johnson, 2004). Zudem konnten Fälle von Kulturheterogenität insbesondere am Beispiel der Stochastizität genregulatorischer Netzwerke immer wieder erklärt werden (Eldar & Elowitz, 2005; McAdams & Arkin, 1997; Elowitz *et al.*, 2002).

Offensichtlich wird die Kulturheterogenität nicht direkt durch die Individuen einer Population reguliert. Es ergibt sich aber die Frage, inwiefern solche Effekte dennoch einen evolutorischen Einfluss auf die *Fitness* einer Art haben. Tatsächlich konnte anhand vieler Beispiele gezeigt werden, dass eine mikrobielle Population durch ihre phänotypische Diversifikation eine höhere Fitness in fluktuierenden Umweltbedingungen hat, als eine phänotypische homogene Population der selben Art haben würde (Avery, 2006; Ingham *et al.*, 2008; McAdams & Arkin, 1999; Veening *et al.*, 2008).

1.4.2 Populationsdynamik

Unter den populationsdynamischen Modellen werden mathematische Methoden zusammengefasst, mit deren Hilfe sich die Fitness von Mischpopulationen beschreiben und untersuchen lassen. Dies geschieht in aller Regel über zeitliche und räumliche Zustandsräume. Eine Gruppe der frühesten Modelle dieser Art¹ stellen die Lotka-Volterra-Gleichungen dar (Lotka, 1934). Dies sind gekoppelte, nicht-lineare Differentialgleichungssysteme, die geeignet sind Räuber-Beute-Dynamiken zu untersuchen. Grundlage dieser Modelle ist, dass die Fitness der einen Subpopulation (z.B. der Räuber) keine ausschließlich inhärente Größe ist, sondern auch von der Größe der jeweils anderen Subpopulation (z.B. der Beute) abhängt. So einfach dieses Modell auch anmuten mag, seine mathematische Analyse liefert interessante Einblicke in verschiedene Aspekte dynamischer Systeme, wie die Periodizität der Populationsgrößen, dem Einfluss von Anfangsbedingungen oder externer Störungen der Populationsgrößen von außen (Murray, 2002).

1.4.3 Replikatordynamik

Eine Generalisierung erfuhr das Lotka-Volterra-Modell durch die evolutionäre Spieltheorie. Um diese in ihrer Eleganz zu erfassen, ist es zunächst hilfreich, sich auf das folgende Gedankenexperiment der klassischen Spieltheorie einzulassen: Zwei Spieler sind mit einer begrenzten Menge von *Strategien* ausgestattet und treffen aufeinander. Jeder Spieler wählt eine der ihm zur Verfügung stehenden Strategien

¹Tatsächlich geht das früheste bekannte populationsdynamische Modell auf Leonardo da Pisa zurück, der in seinem Werk *Liber Abbaci* die dynamische Entwicklung von Kaninchenpopulationen beschrieb. Die Gleichung $x_n = x_{n-1} + x_{n-2}$; $n \geq 2$, die er hierzu aufstellte war nicht nur das erste dynamische Modell überhaupt, sondern hat die Fibonacci-Zahlenfolge zum Ergebnis, die sich als Randnotiz in dem entsprechenden Kapitel findet.

ohne vom jeweils anderen zu wissen, welche Strategie dieser wählt. Jede Kombination der beiden Strategien führt zu einer bestimmten (u.U. negativen) Auszahlung an beide Spieler. Solche Spiele lassen sich in der sog. *Normalform* darstellen, wozu die folgenden Definitionen benötigt werden:




Gegeben sind Spieler \mathcal{F} , der aus m verschiedenen Strategien wählen kann und \mathcal{G} , ein weiterer Spieler, der aus n Strategien wählen kann. Weiterhin gibt der Koeffizient $f_{i,j}$ die Auszahlung an Spieler \mathcal{F} an, wenn dieser Strategie i und Spieler \mathcal{G} Strategie j wählt. Alle Koeffizienten lassen sich in der sog. *Auszahlungsmatrix* $\mathbf{F} \in \mathbb{R}^{m \times n}$ anordnen. Letztendlich sei ein Vektor $\mathbf{p} \in \mathbb{R}^m$, für den $p_i = 1$ ist, wenn Spieler \mathcal{F} Strategie i wählt und an allen anderen Stellen 0. Analog sei $\mathbf{q} \in \mathbb{R}^n$ ein Vektor mit $q_j = 1$ der die gewählte Strategie j von Spieler \mathcal{G} ausdrückt. Sei $u(\mathbf{p}, \mathbf{q})$ die Auszahlung des konkreten Spiels. Damit lässt sich das gesamte Spiel als algebraische Gleichung darstellen:

$$u(\mathbf{p}, \mathbf{q}) = \mathbf{p}^T \cdot \mathbf{F} \cdot \mathbf{q} \quad (1.1)$$

Analog lassen sich mit $\mathbf{G} \in \mathbb{R}^{n \times m}$ die Auszahlungen an Spieler \mathcal{G} ausdrücken. Die Auszahlung $v(\mathbf{q}, \mathbf{p})$ seines konkreten Spiels ergibt sich dann mit:

$$v(\mathbf{q}, \mathbf{p}) = \mathbf{q}^T \cdot \mathbf{G} \cdot \mathbf{p} \quad (1.2)$$

Ein Spiel heißt symmetrisch, wenn die beiden Spieler \mathcal{F} und \mathcal{G} dadurch ununterscheidbar sind, indem sie über die selben Strategien verfügen und die Auszahlungsmatrices identisch sind: $\mathbf{F} = \mathbf{G}$.

Ein Beispiel soll dieses Gedankenexperiment veranschaulichen: \mathcal{F} und \mathcal{G} spielen das Pausenhofspiel Schere-Stein-Papier: Schere () schlägt Papier wird aber von Stein geschlagen, Stein () schlägt Schere wird aber von Papier geschlagen, Papier () schlägt Stein, wird aber von Schere geschlagen. Die Auszahlung für den Gewinner

soll 1, für den Verlierer -1 betragen. \mathcal{F} entscheidet sich für Schere, \mathcal{G} für Stein. Aus Sicht von \mathcal{F} stellt sich das Spiel entsprechend dar:

$$u\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right) = (1, 0, 0) \cdot \left(\begin{array}{c|ccc} & \text{✂} & \text{○} & \text{📄} \\ \hline \text{✂} & 0 & -1 & 1 \\ \text{○} & 1 & 0 & -1 \\ \text{📄} & -1 & 1 & 0 \end{array} \right) \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -1$$

Offensichtlich verliert \mathcal{F} , während \mathcal{G} durch die Auszahlung $v\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = 1$ gewinnt.

Entscheidet sich ein Spieler nur für eine Strategie wie im obigen Beispiel, spricht man von *reinen* Strategien. Wenn allerdings jeder Spieler eine Strategie mit einer bestimmten Wahrscheinlichkeit wählt (bei z.B. hinreichend vielen Wiederholungen des Spiels), spricht man von einer *gemischten* Strategie. Die Wahrscheinlichkeitsverteilung wird dann entsprechend wieder durch \mathbf{p} und \mathbf{q} mit $\sum_i p_i = 1$ und $\sum_i q_i = 1$ ausgedrückt. 1950 fand Nash, dass es in Spielen mit gemischten Strategien für \mathcal{F} immer mindestens eine Strategieverteilung \mathbf{p}^* gibt, für die $u(\mathbf{p}^*, \mathbf{q}) \geq u(\mathbf{p}, \mathbf{q})$ mit $\mathbf{p}^* \neq \mathbf{p}$ gilt; es also eine gemischte Strategie gibt, die über hinreichend viele Spiele als beste Antwort auf die gemischte Strategie \mathbf{q} gilt. Unter der Annahme, dass \mathcal{A} daher diese gemischte Strategie spielen wird, gibt es analog auch eine beste Antwort \mathbf{q}^* von der \mathcal{G} nicht unilateral abweichen kann ohne seine durchschnittliche Auszahlung zu reduzieren. Die Existenz des sog. *Nash-Gleichgewichts* $(\mathbf{p}^*, \mathbf{q}^*)$ hatte weitreichende Konsequenzen für Wirtschafts-, Sozial- und Politikwissenschaften und letztendlich durch die Entwicklung der evolutionären Spieltheorie (ES) auch in der Ökologie (Smith, 1974, 1976).

Überträgt man die klassische Spieltheorie auf ökologische Mischpopulationen, stellt jede Population einen Spieler dar. Jedes Individuum einer Population *wählt* eine aus der ihm zur Verfügung stehen-

den Strategien, wobei die *Wahl* meist phäno- oder genotypisch determiniert ist. Die Auszahlungen werden in Form positiver oder negativer Reproduktionsraten ausgedrückt. Im Folgenden werden nun nur noch symmetrische Spiele betrachtet. Damit reduziert sich bezogen auf den Populationsbegriff die Anzahl der betrachteten Populationen auf eine, mit der Auszahlungsmatrix \mathbf{F} und der Populationsverteilung \mathbf{x} als gemischter Strategie. Die Nettoreproduktionsrate der Gesamtpopulation ist somit

$$\mu = \mathbf{x}^T \cdot \mathbf{F} \cdot \mathbf{x}.$$

Die Reproduktionsrate der Subpopulation mit der i -ten Strategie ist demnach

$$\mu_i = (\mathbf{F} \cdot \mathbf{x})_i.$$

Interessiert man sich nun nur für die relativen Änderungen der Subpopulationen untereinander, ist es hinreichend die Differenzen der Reproduktionsraten in ein dynamisches Modell zu integrieren. Dies wird durch die Replikatorgleichung ausgedrückt (Schuster & Sigmund, 1983):

$$\frac{dx_i(t)}{dt} = x_i \cdot ((\mathbf{F}\mathbf{x})_i - \mathbf{x}^T \mathbf{F} \mathbf{x}) \quad (1.3)$$

Dieses Modell wirkt deswegen so attraktiv, weil es die Untersuchung der Existenz stabiler Zustände, sog. Fixpunkte, \mathbf{x}^* zulässt. Da das Nash-Gleichgewicht eine *Beste-Antwort-Dynamik* darstellt, ist es per se ein solcher Fixpunkt. Eine Population deren Verteilung der Subpopulationen \mathbf{x}^* erreicht, wird sich daher nicht mehr ändern. Eine Erweiterung dessen sind evolutionär stabile Zustände (ESS). Ein ESS ist ein Nash-Gleichgewicht, das um ein hinreichend kleines $\Delta \mathbf{x}$ gestört, im zeitlichen Verlauf wieder auf \mathbf{x}^* zustreben wird. Letztendlich unterscheidet man *innere* und *äußere* ESS. Ein

äußerer ESS hat für eine oder mehrere Strategien (also Subpopulationen) $x_i^* = 0$. Diese Subpopulationen werden also im Verlauf der Spieldynamik extingieren. Für einen inneren ESS gilt hingegen $x_i^* \neq 0$ für alle Strategien. In symmetrischen Spielen haben nach dem Satz von Bishop & Cannings (1978) innere ESS zuweilen die Eigenschaft immer der einzig existierende ESS zu sein.

Weitergehende Einführung in die Thematik finden sich in Cressman (2003); Hofbauer & Sigmund (2003); Sandholm (2010); Szabó & Fáth (2007) oder Novak (2006).

1.5 Zielsetzung

Ziel dieser Arbeit war es, Modelle und Methoden zu entwickeln und zu validieren, die physiologische Prozesse in mikrobiologischen Populationen beschreiben und quantitativ erfassen können.

In einem ersten Projekt sollte hierzu ein neuartiger Analyseprozess zur Auswertung von Hochdurchsatz-Metabolomdaten entwickelt, bioinformatisch in Form einer Software implementiert und validiert werden. Ziel des Prozesses sollte es sein, diejenigen Metabolite und Stoffwechselwege zu identifizieren, die über verschiedene Umweltbedingungen oder Zustände mikrobiologischer Populationen variiert werden. Hierzu sollten die Daten zunächst durch unüberwachte Clusterverfahren entsprechend ihrer Ähnlichkeit gruppiert werden. Signifikante Variationen der Metabolite und Stoffwechselwege ließen sich dann mittels Methoden der schließenden Statistik herausfiltern. Eine Visualisierung der Ergebnisse sollte die zugrundeliegenden biologischen Zusammenhänge intuitiv erfassbar machen. Eine quantitative Gegenüberstellung der implementierten Clustermethoden sollte Klarheit über den Nutzen der ESOMs in der Metabolomik schaffen.

In einem zweiten Projekt sollte gezielt die Dynamik der BChl-*a* Produktion in *D. shibae* Populationen unter fluktuierenden Umweltbedingungen untersucht werden (*Roseobacter*-Projekt). Hierzu sollte zunächst ein spieltheoretisches Modell populationsdynamisch erklären, ob die zur Licht-Exposition azyklische BChl-*a* Produktion für die Fitness der Population vorteilhaft ist.

Die Ergebnisse dieser theoretischen Analyse des zweiten Projektes sollten anschließend experimentell verifiziert werden. Hierzu sollten zunächst neue instrumentelle Methoden der Batch-Mikroreaktionstechnik entwickelt werden. Ziel dieser neuen Methoden sollte sein, im Hochdurchsatz relevante Prozessparameter abschätzbar zu machen.

Der Fokus bei der Entwicklung dieser Methoden lag auf der Vermeidung ungewollter Einflüsse durch experimentelle Handhabung und gleichzeitig einer Vervielfachung der Grundgesamtheit der Ergebnisse durch eine Erhöhung der parallelen Ansätze. Ferner musste in dem Zusammenhang ein Verfahren entwickelt werden, um die Menge an produzierten BChl-*a* interaktionsfrei *in-vivo* quantifizieren zu können.

2 Materialien und Methoden

2.1 Methoden der Software-Entwicklung

2.1.1 Die Entwicklungswerkzeuge

2.1.1.1 Programmierung in Java

Die Software eSOMet wurde ausschließlich in der Programmiersprache Java¹ entwickelt. Der Programmcode wurde gegen die Version 1.6 kompiliert. Als Entwicklungswerkzeug kam das Java Development Kit (JDK) 5 bzw. JDK 6 und als Entwicklungsumgebung die Eclipse-Plattform² in der jeweils aktuellen Version zum Einsatz.

2.1.1.2 Programmierung in C#

Zur Steuerung der Automatisierungsschnittstelle des Gerätes VarioskanFlash wurde ein Konsolenprogramm in der Programmiersprache C#³ entwickelt und gegen die Version 3.5 kompiliert. Als Entwicklungswerkzeug kam .NET 3.5⁴ und als Entwicklungsumgebung Microsoft VisualStudio⁵ zum Einsatz.

¹<http://www.oracle.com/technetwork/java>

²<http://www.eclipse.org>

³<http://msdn.microsoft.com/de-de/vcsharp/>

⁴<http://www.microsoft.com/net>

⁵<http://www.microsoft.com/VisualStudio2010>

2.1.1.3 Programmierung und Datenauswertung in R

Zur Auswertung und Visualisierung experimenteller Daten, sowie zur numerischen Integration von Differentialgleichungssystemen wurde die Skriptsprache R⁶ eingesetzt. Die erzeugten R-Skripte wurden durch den R-Interpreter in den Versionen 2.8 bzw. 2.9 ausgeführt.

2.1.1.4 Verwendete Bibliotheken

Im Rahmen dieser Arbeit wurden externe Programmierbibliotheken verwendet, die nicht im Standardumfang der jeweiligen Programmierbibliotheken waren (Tabelle 2.1).

2.1.2 Die KEGG Datenbank

Zur Interpretation statistischer Auswertungen von Metabolomdaten war es notwendig, die metabolischen Profile in einen biologisch relevanten Zusammenhang zu setzen. Die hierzu notwendigen Informationen wurden der, auf Stoffwechselprozesse spezialisierten, Datenbank Kyoto Encyclopedia of Genes and Genomes (KEGG)⁷ entnommen (Kanehisa & Goto, 2000; Kanehisa *et al.*, 2006, 2008). Diese stellt derzeit die größte, frei verfügbare Plattform solcher Daten dar und ist in verschiedene Sektionen eingeteilt. Innerhalb dieser waren für diese Arbeit die Sektionen KEGG Pathway und KEGG Ligand von Relevanz.

Die entsprechenden Informationen ließen sich als Klartextdateien aus dem FTP- (*engl.* File Transfer Protocol) Bereich der KEGG Datenbank⁸ herunterladen und durch entsprechend programmierte

⁶<http://www.r-project.org>

⁷<http://www.genome.jp/kegg>

⁸<ftp://ftp.genome.jp/pub/kegg/>

Tabelle 2.1: In dieser Arbeit verwendete Programmierbibliotheken.

Bibliothek	Sprache	URL, Referenz
metabolomsom	Java	(Hiller, 2006)
kohonen	Java	(Hiller, 2006)
Classifier4J	Java	http://classifier4j.sourceforge.net
The Colt Project	Java	http://acs.lbl.gov/hoschek/colt
Apache Commons	Java	http://commons.apache.org
Flanagan Package	Java	http://www.ee.ucl.ac.uk/~mflanaga/java
HSQldb	Java	http://hsqldb.org
iText	Java	http://www.lowagie.com/iText
JDOM	Java	http://jdom.org
JFreeChart	Java	http://www.jfree.org/jfreechart
HidLibrary	C#	http://github.com/mikeobrien/HidLibrary
SkanIt AI	C#	Thermo Fisher Scientific Inc., Waltham, USA
deSolve	R	(Soetaert <i>et al.</i> , 2010)
ggplot2	R	(Wickham, 2009)

Parser auslesen. Um die Informationen wiederum effizient verfügbar zu machen, wurden alle heruntergeladenen und geparsen Daten in der SQL (*engl.* Structured Query Language) Datenbank HSQldb reorganisiert.

2.1.3 Biostatistische Methoden

2.1.3.1 Multivariate Datenmatrix

Als erstes Ergebnis einer semi-quantitativen, gaschromatographisch-massenspektrometrischen (GC-MS) Analyse erhielt man für die gemessenen chemischen Komponenten verhältnisskalierte Daten. Diese wurden in einer *multivariaten Datenmatrix* $\mathbf{X} \in \mathbb{R}^{p \times q}$ organisiert,

so dass p der Anzahl der *Variablen* (chemische Komponenten) und q der Anzahl der *Beobachtungen* (durchgeführten Experimente, Wiederholungen) entsprach. Alle Werte einer Beobachtung j wurden in dem jeweiligen Spaltenvektor \mathbf{X} : $\mathbf{s}_j = (x_{1,j}, x_{2,j}, \dots, x_{p,j})^T$, alle Werte einer Variablen i in dem jeweiligen Zeilenvektor \mathbf{X} : $\mathbf{c}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,q})$ der Matrix zusammengefasst.⁹

2.1.3.2 Datennormalisierung

Multivariate Daten unterliegen immer einer Streuung innerhalb einer Variablen. Insbesondere die Rohdaten eines GC-MS-Experimentes können dabei eine Streuung über Größenordnungen aufweisen, die unter Umständen nicht mehr biologisch relevant sind (van den Berg *et al.*, 2006). Ferner erfordern die Verfahren der multivariaten Statistik immer eine bestimmte stochastische Verteilung der jeweiligen Variablen. Hierzu erzeugen Normalisierungsverfahren die jeweils skalierten Matrizes $\mathbf{N}_\bullet \in \mathbb{R}^{p \times q}$. Sechs solcher Verfahren wurden implementiert:

Z-Transformation erzeugte die Matrix \mathbf{N}_Z mit den Koeffizienten

$$n_{i,j} = \frac{x_{i,j} - \bar{x}_i}{s_i} \text{ mit } \bar{x}_i \text{ als dem Mittelwert und } s_i \text{ als der Standardabweichung aller } \mathbf{c}_i.$$

Intervall Skalierung erzeugte die Matrix \mathbf{N}_R mit den Koeffizienten

$$n_{i,j} = \frac{x_{i,j} - \bar{x}_i}{x_{i_{\max}} - x_{i_{\min}}} \text{ mit } x_{i_{\max}} \text{ als dem Maximum und } x_{i_{\min}} \text{ als dem Minimum aller } \mathbf{c}_i.$$

Pareto Skalierung erzeugte die Matrix \mathbf{N}_P mit den Koeffizienten

$$n_{i,j} = \frac{x_{i,j} - \bar{x}_i}{\sqrt{s_i}}.$$

⁹Die hier gewählte Notation weicht von der typischen Notation multivariater Matrizes ab, in welchen normalerweise die Spaltenvektoren die Variablen und die Zeilenvektoren die Beobachtungen repräsentieren.

Breiten Skalierung erzeugte die Matrix \mathbf{N}_V mit den Koeffizienten

$$n_{i,j} = \frac{x_{i,j} - \bar{x}_i}{s_i} \cdot \frac{\bar{x}_i}{s_i}.$$

Niveau Skalierung erzeugte die Matrix \mathbf{N}_L mit den Koeffizienten

$$n_{i,j} = \frac{x_{i,j} - \bar{x}_i}{\bar{x}_i}.$$

Gesamtsignal Skalierung skalierte die Daten einer Beobachtung zu deren relativen, kumulierten Gesamtsignal (Villas-Bôas *et al.*, 2005). Sie erzeugte die Matrix \mathbf{N}_S mit den Koeffizienten $n_{i,j} =$

$$\frac{x_{i,j}}{\lambda_j} \text{ mit } \lambda_j = \frac{\sum_{i=1}^q x_{i,j}}{\frac{1}{q} \sum_{l=1}^q \sum_{k=1}^p x_{k,l}}.$$

2.1.3.3 Hauptkomponentenanalyse

Die Hauptkomponentenanalyse (PCA) ist ein Verfahren, welches multivariate Daten auf einen neuen Vektorraum projiziert (Villas-Bôas *et al.*, 2007). Die neuen Basisvektoren sind dabei weiterhin orthonormal zu einander und *erklären* in fallender Reihenfolge die größten Varianzen innerhalb der Daten. Die Eingangsdaten \mathbf{N}_Z müssen dazu standardnormalverteilt, also Z-transformiert sein. Aus den normalisierten Daten ließ sich die Kovarianz-Matrix $\Sigma \in \mathbb{R}^{p \times p}$ berechnen

$$\Sigma = \frac{1}{q-1} \mathbf{N}_Z \mathbf{N}_Z^T \quad (2.1)$$

Auf dieser wurde die Eigenwertzerlegung

$$\Sigma = \mathbf{A} \mathbf{\Lambda} \mathbf{A}^T \quad (2.2)$$

durchgeführt, wobei jeder Spaltenvektor aus $\mathbf{A} \in \mathbb{R}^{p \times p}$ einen Eigenwertvektor zu Σ und $\mathbf{\Lambda}$ eine Diagonalmatrix mit den korrespondierenden Eigenwerten repräsentierte. Die Koeffizienten in $\mathbf{\Lambda}$ wurden so neusortiert, dass $\lambda_{1,1} \geq \lambda_{2,2} \geq \dots \geq \lambda_{p,p}$. Die entsprechend neusortierten Spaltenvektoren aus \mathbf{A} definierten eine neue Basis des ur-

sprünglichen Vektorraums und werden Hauptkomponenten genannt. Die Projektion \mathbf{X}^{PCA} der Daten wurde durch

$$\mathbf{X}^{\text{PCA}} = \mathbf{A}^T \mathbf{N}_Z \quad (2.3)$$

berechnet.

2.1.3.4 Hierarchische Clusteranalyse

Die hierarchische Clusteranalyse (HCA) ist ein *agglomeratives* Clusterverfahren für multivariate Daten. Es fasst die Beobachtungen hierarchisch derart zu disjunkten Clustern zusammen, dass solche Beobachtungen, die sich besonders ähnlich sind, in gleichen Clustern gruppiert werden. Als Maß der Ähnlichkeit wurde die Distanz

$$d_{k,l} = \|\mathbf{s}_l - \mathbf{s}_k\|^n, \quad (2.4)$$

zweier Beobachtungsvektoren \mathbf{s}_k and \mathbf{s}_l gewählt, wobei $\|\cdot\|^n$, $n \in [1, 2, \dots, \infty)$ eine Minkowski-Metrik erzeugende n -Norm definiert:

$$\|\mathbf{s}_l - \mathbf{s}_k\|^n = \left(\sum_{i=1}^p |s_{i,l} - s_{i,k}|^n \right)^{\frac{1}{n}} \quad (2.5)$$

Um die Distanz zweier disjunkter Cluster $\alpha = \{\mathbf{s}_t, \dots, \mathbf{s}_u\}$ und $\beta = \{\mathbf{s}_v, \dots, \mathbf{s}_w\}$ höherer Hierarchie zu bestimmen, wurden drei Distanzfunktionen mit jeweils $\mathbf{s}_i \in \alpha$ und $\mathbf{s}_j \in \beta$ implementiert:

Nächstgelegener Nachbar (Single Linkage) ist als die kleinste Distanz aller paarweisen Beobachtungen beider Cluster definiert:

$$\delta(\alpha, \beta) = \min(d_{i,j}).$$

Entferntester Nachbar (Complete Linkage) ist als die größte Distanz aller paarweisen Beobachtungen beider Cluster definiert:

$$\delta(\alpha, \beta) = \max(d_{i,j}).$$

Durchschnittlich entfernter Nachbar (Average Linkage) ist als die durchschnittliche Distanz aller paarweisen Beobachtungen beider Cluster definiert: $\delta(\alpha, \beta) = \frac{\sum_i \sum_j d_{i,j}}{|\alpha| \cdot |\beta|}$.

Algorithmus 1 beschreibt das Clustern, welches als Ergebnis eine binäre Baumstruktur (auch Dendrogramm) liefert, im Detail. Das Dendrogramm bildet die gesamte Ähnlichkeitsstruktur der Daten ab und gibt sie durch eine weitergehende Visualisierung intuitiv wieder.

Algorithmus 1: Hierarchische Clusteranalyse

```
Input :
  Multivariate Datenmatrix X
  Ordnung  $n$  der  $n$ -Norm
  Distanzfunktion  $\delta$ 
  Normalisierungsfunktion  $\nu$ 
Output :
  Projektion aller  $\mathbf{c}_j \in \mathbf{X}$  auf den Binärbaum  $\Gamma$ 
begin
   $\mathbf{N} \leftarrow \nu(\mathbf{X})$ ;
   $\mathbf{D}^n = \mathbf{0}^{p \times q}$ ;
  // Berechnung der Distanzmatrix aus den Spaltenvektoren  $\mathbf{s}$  aus  $\mathbf{N}$ .
  for  $k \in [1 \dots q - 1]$  do
    for  $l \in [k + 1 \dots q]$  do
       $\mathbf{D}_{k,l}^n = d_{k,l} = \|\mathbf{s}_l - \mathbf{s}_k\|^n$ ;
  // Initialisierung des Dendrogrammes und seiner Blätter.
   $\Gamma \leftarrow \emptyset$ ;
  for  $k \in [1 \dots q]$  do
     $\gamma_k \leftarrow \{\mathbf{s}_k\}$ ;
     $\Gamma \leftarrow \Gamma \cup \{\gamma_k\}$ ;
  // Eigentlicher Clusterprozess.
```

2.1.3.5 Selbstorganisierende Karten

Emergente selbstorganisierende Karten (ESOM) sind ein Werkzeug der künstlichen Intelligenz und eine spezielle Anwendung künstlicher neuronaler Netze. Sie werden dazu genutzt, in multivariaten Daten nach Mustern zu suchen und diese entsprechend ihrer Ähn-

lichkeit topologisch zu sortieren. Hierzu wurden die Beobachtungsvektoren \mathbf{s}_j (auch Eingangsvektoren) auf einem zweidimensionalen Gitter \mathbf{M} , bestehend aus $X \times Y$ künstlichen Neuronen $(M_{x,y})$, angeordnet. Emergente SOMs unterscheiden sich von Gewöhnlichen insoweit, als dass die Anzahl der Neuronen mindestens eine Größenordnung höher ist als die der Eingangsvektoren. Um unbeabsichtigte Randeffekte zu verhindern, lässt sich dieses Gitter in einen randlosen, finiten Torus einbetten. Alle Knoten der Eingabeschicht sind mit den Knoten des Gitters durch die Gewichtsvektoren $\mathbf{w}_{x,y} \in \mathbb{R}^p$ verbunden.

Während des unbeaufsichtigten Lernprozesses der ESOMs wurden die Gewichtsvektoren stets so angepasst, dass sie die der Eingangsdaten hochdimensional zugrundeliegende Struktur auf das zweidimensionale Gitter abbildeten. Als Ergebnis des Trainingsprozesses (Algorithmus 2) wurden solche Eingangsvektoren auf dem Gitter in räumlicher Nähe angeordnet, deren Muster sich ähnlich sind (Kohonen, 1982).

Aufgrund der hohen Anzahl der verwandten Neuronen wurden die Eingangsvektoren zerstreut auf dem Ausgabegitter verteilt. Um die den Daten zugrundeliegende Struktur aus der trainierten ESOM zu erfassen, wurde die U*-Matrix $\mathbf{U}^* \in \mathbb{R}^{x \times y}$ berechnet (Ultsch, 1999). Hierzu wurde für jedes Neuron $M_{x,y}$ die Summe der euklidischen Distanzen zwischen seinen Gewichtsvektoren und denen seiner vier unmittelbaren Nachbarn berechnet. Die U*-Matrix wurden als dreidimensionale Grafik visualisiert und unterteilt die Ausgabeschicht in verschiedene Bereiche: Eingangsvektoren, die eine hohe Ähnlichkeit haben, wurden in gemeinsamen „Tälern“ der drei-dimensionalen Landschaft eingebettet, wobei solche Eingangsvektoren mit hoher Distanz durch „Berge“ separiert werden.

Algorithmus 2: Training selbstorganisierender Karten aus multivariaten Daten

Input :

Normalisierte, multivariate Datenmatrix \mathbf{N}
 Anzahl der Trainingsschritte t_{\max}
 Trainingsrate $\alpha \in [0, 1]$
 Initialer Trainingsradius ν
 Größe der Ausgabeschicht $X \times Y$

Output :

Die Menge aller trainierten Gewichtsvektoren $\mathbf{w}_{x,y} \in \mathbf{M}$

```

begin
  // Initialisierung des Gitters.
  for  $x \in [1 \dots X]$  do
    for  $y \in [1 \dots Y]$  do
      // Zufällige Initialisierung der Gewichtsvektoren.
      for  $k \in [1 \dots p]$  do
         $(\mathbf{w}_{x,y})_k \leftarrow \text{RANDOM}()$ ;

  // Beginn des Trainings
  for  $t \in [1 \dots t_{\max}]$  do
     $\mathcal{V} \leftarrow \{1, \dots, q\}$ ;
    // Aktualisierung der Trainingsrate.
     $\tau(t) \leftarrow \alpha \cdot \left(1 - \frac{t-1}{t_{\max}}\right)$ ;
    // Aktualisierung des Trainingsradius.
     $\sigma(t) \leftarrow \frac{1}{2} \tau(t) \cdot \nu \cdot \sqrt{X^2 + Y^2}$ ;
    while  $\mathcal{V} \neq \emptyset$  do
      // Zufällige Wahl eines Spaltenvektors  $\mathbf{s}_j$  aus  $\mathbf{N}$ .
       $j \leftarrow \text{RANDOM}(\mathcal{V})$ ;
       $\mathcal{V} \leftarrow \mathcal{V} \setminus k$ ;
      // Bestimmung des Gewichtsvektors  $\mathbf{w}_{x^*,y^*}$  mit der höchsten
      Ähnlichkeit zum gewählten Spaltenvektor  $\mathbf{s}_j$ .
       $d \leftarrow \infty$ ;
       $\{x^*, y^*\} \leftarrow \{\}$ ;
      for  $k \in [1, \dots, X]$  do
        for  $l \in [1, \dots, Y]$  do
          if  $\|\mathbf{s}_j - \mathbf{w}_{k,l}\|^2 < d$  then
             $\{x^*, y^*\} \leftarrow \{k, l\}$ ;

      // Anpassung des dichtesten Gewichtsvektors.
       $\mathbf{w}_{x^*,y^*} \leftarrow \mathbf{w}_{x^*,y^*} + \tau(t) \cdot (\mathbf{s}_j - \mathbf{w}_{x^*,y^*})$ ;
      // Anpassung aller in der Umgebung liegenden Gewichtsvektoren.
      for  $k \in [1, \dots, X]$  do
        for  $l \in [1, \dots, Y]$  do
          // Bestimmung der kürzesten Entfernung  $r$  der beiden
          Neuronen  $M_{k,l}$  und  $M_{x^*,y^*}$  auf dem Gitter. Die toroide
          Einbettung des Gitters ist hierbei berücksichtigt.
           $r \leftarrow \left\| \begin{pmatrix} x^* \\ y^* \end{pmatrix} - \begin{pmatrix} k \\ l \end{pmatrix} \right\|^2$ ;
          // Bestimmung der Gauss'schen Trainingsrate  $\eta(t, r)$ .
           $\eta(t, r) \leftarrow e^{-\frac{1}{2} \left(\frac{r}{\sigma(t)}\right)^2}$ ;
           $\mathbf{w}_{k,l} \leftarrow \mathbf{w}_{k,l} + \tau(t) \cdot \eta(t, r) \cdot (\mathbf{s}_j - \mathbf{w}_{k,l})$ ;

```

2.1.3.6 Dendrogramme

Die Ergebnisse des hierarchischen Clusters, wie auch der Anwendung der ESOMs ließen sich durch *Dendrogramme* visualisieren. Im Fall der HCA ergab sich das relative Niveau, auf dem jeweils zwei Cluster verbunden wurden, aus der berechneten Distanz zweier Cluster. Im Fall der ESOMs ließ sich dieses Niveau durch ein „virtuelles Fluten“ der zugehörigen U^* -Matrix-Landschaft bestimmen: Das Niveau auf dem zwei Täler durch das Fluten miteinander verbunden worden waren entsprach der Höhe, zu der auch die Cluster im Dendrogramm miteinander verbunden wurden (Ultsch, 2003).

2.1.3.7 Silhouettenbreite

Die Qualität der berechneten Cluster ließ sich durch die durchschnittliche *Silhouetten-Breite* eines Clusters (Rousseeuw, 1987) messen. Sie beruhte auf dem Vergleich der Inter-Clusterdistanzen zu den Intra-Clusterdistanzen eines Beobachtungsvektors zu seinen Nachbarn. Die Silhouettenbreite $S(j)$ einer Beobachtung j in einem Cluster α berechnete sich aus

$$S(j) = \frac{b(j) - a(j)}{\max((a(j), b(j)))} \quad (2.6)$$

mit $b(j)$ als der durchschnittlichen euklidischen Distanz des Beobachtungsvektors \mathbf{s}_j zu allen Beobachtungsvektoren im Nachbarcluster β und $a(j)$ als der durchschnittlichen euklidischen Distanz des Beobachtungsvektors \mathbf{s}_j zu allen Beobachtungsvektoren im eigenen Cluster α . Offensichtlich gilt immer $S(j) \in [-1; 1]$. Eine Silhouettenbreite von ungefähr 1 bedeutete, dass $a(j)$ sehr klein im Vergleich zu $b(j)$ ist und j dem richtigen Cluster zugeordnet wurde. Umgekehrt deutete ein sehr kleiner Wert von $S(j)$ auf eine Fehlzusammenordnung hin. Im Falle von $S(j) \approx 0$ ließ sich keine eindeutige Zuordnung von j zu α oder β vornehmen.

Die durchschnittliche Silhouttenbreite eines Clusters γ wurde definiert als

$$S_\gamma = \frac{1}{|\gamma|} \cdot \sum_k S(k) \quad (2.7)$$

mit $k \in \gamma$ und der Anzahl aller Elemente $|\gamma|$ in γ .

2.1.3.8 Tests der schließenden Statistik

Zur Identifikation der zwischen den Beobachtungen zweier Cluster signifikant variierten Metabolitabundanzanzen wurde der *Student t-Test* (Press *et al.*, 1992) angewandt, nachdem die geclusterten Daten einer Gesamtsignal-Skalierung unterzogen wurden. Das Signifikanzniveau wurde mit $p < 0,05$ festgelegt.

Um durch die Menge der signifikant variierten Metabolite diejenigen biochemischen Reaktionsnetzwerke zu identifizieren, die am stärksten beeinflusst waren, wurden die Daten der KEGG Datenbank verwendet. Auf Basis dieser Informationen wurde für jedes Reaktionsnetzwerk eine Kontingenztafel erzeugt und mit Hilfe des exakten Tests nach Fisher die Wahrscheinlichkeit bestimmt, mit der ein Reaktionsnetzwerk zufällig beeinflusst war.

2.2 Experimentelle Methoden

2.2.1 Geräte

Im Rahmen experimenteller Arbeiten wurden die in Tabelle 2.2 aufgeführten Geräte verwendet.

Tabelle 2.2: In dieser Arbeit verwendeten Geräte.

Gerät	Modell / Bezeichnung	Referenz
Autoklav	FVA/A1	Fedegari AG, Bedano, Schweiz
Fluoreszenz- mikroskop	Axiovert 200, N XBO 75, Axio- Cam b/w	Carl Zeiss MicroImaging GmbH, Jena, Deutsch- land
Multi-Label Lesege- rät	Varioskan Flash	Thermo Fisher Scientific Inc., Waltham, USA
pH-Messgerät	pH211 Mikro- prozessor	pH-Meter HANNA In- struments, Woonsocket, USA
Photometer	Ultrospec2100pro	GE Healthcare Europe, München, Deutschland
Pipetten	Reference	Eppendorf, Hamburg, Deutschland
Reinstwasseranlage Schüttler	Milli-Q System TR150	Millipore, Billerica, USA Infors, Bottmingen, Schweiz
Sterilarbeitsbank	HERAsafe	Heraeus Holding GmbH, Hanau, Deutschland

Fortsetzung der vorhergehenden Seite

Gerät	Modell / Bezeichnung	Referenz
Waagen	ACL-3100.2, BL1500, BL61S	Sartorius, Göttingen, Deutschland

2.2.2 Chemikalien und Verbrauchsmaterialien

Soweit nicht anders angegeben wurden Chemikalien, Reagenzien und Verbrauchsmaterialien von folgenden Firmen bezogen: Fluka (Buchs, Schweiz), Merck (Darmstadt, Deutschland), Millipore (Billerica, USA), Qiagen (Hilden, Deutschland), Riedel-de Häen (München, Deutschland), Roche (Grenzach-Wyhlen, Deutschland), Roth (Karlsruhe, Deutschland), Nunc (Langensbold, Deutschland) und Sigma-Aldrich (St. Louis, USA).

2.2.3 Kultivierungsbedingungen von *D. shibae*

2.2.3.1 Bakterienstamm

Im Rahmen dieser Arbeit wurde mit dem Bakterienstamm *D. shibae* DFL12^T (Biebl *et al.*, 2005; Wagner-Döbler *et al.*, 2010) gearbeitet, der ursprünglich von Zellen des marinen Dinoflagellaten *Prorocentrum lima* isoliert wurde.

2.2.3.2 Medien

Die Kultivierung von *D. shibae* erfolgte entweder auf bzw. in Marine Bouillon (MB)-Medium oder Salzwassermedium (SWM). Alle Puffer und Medien bzw. Medienbestandteile wurden, wenn nicht anders angegeben, im Autoklaven 20 min bei 121 °C und 1 bar Überdruck dampfsterilisiert. Hitzeempfindliche Substanzen wurden ste-

rilfiltriert. Für die Kultivierung auf festem Medium wurden der jeweiligen Lösung 1,5 % (w/v) Agar-Agar zugegeben. Agar-haltiges Medium wurde vor der Sterilisation kurz bis zum Sieden erhitzt, um ein Ausfällen von Medienbestandteilen aufgrund hoher Salzkonzentrationen zu vermeiden.

MB-Medium

Marine Bouillon, (Roth, Karlsruhe, Deutschland) in deion. H ₂ O	40,1 g/L
--	----------

SWM-Basis 5x

H ₂ SO ₄	20,0 g/L
NH ₄ Cl	1,25 g/L
NaCl	100,0 g/L
MgCl ₂ · 6 H ₂ O	15,0 g/L
KCl	2,50 g/L
in deion. H ₂ O	
Einstellung des pH-Wertes auf 7,0 mit NaOH	

CaCl₂ 100x

CaCl ₂ · 2 H ₂ O	10,00 g/L
in deion. H ₂ O	

KH₂PO₄ 100x

KH ₂ PO ₄	20,0 g/L
in deion. H ₂ O	

NaHCO₃ 10x

NaHCO ₃	1,90 g/L
in deion. H ₂ O	

Succinat 10x

Succinat	20,0 g/L
in deion. H ₂ O	

Vitamine 100x

Biotin	0,20 g/L
Nikotinsäure	2,0 g/L
4-Aminobenzoessäure	0,80 g/L
in deion. H ₂ O, anschließend sterilfiltriert	

Spurenelemente 500x

deion. H ₂ O	25 mL
Fe(II)SO ₄ · 7 H ₂ O	1,05 g
HCl 25%ig	6,5 mL
Titriplex (Na ₂ EDTA)	2,6 g
Einstellung des pH-Wertes auf 6,0 - 6,5 mit NaOH	
H ₃ BO ₃	0,015 g
MnCl ₂ · 4 H ₂ O	0,05 g
CoCl ₂ · 6 H ₂ O	0,095 g
NiCl ₂ · 6 H ₂ O	0,012 g
CuCl ₂ · 2 H ₂ O	0,001 g
ZnSO ₄ · 7 H ₂ O	0,072 g
Na ₂ MoO ₄ · 2 H ₂ O	0,018 g
Auffüllen auf 500 ml deion. H ₂ O, anschließend sterilfiltriert	

SWM 1x

SWM-Basis 5x	200 mL
NaHCO ₃ 10x	100 mL
Succinat 10x	100 mL

KH ₂ PO ₄ 100x	10 mL
CaCl ₂ 100x	10 mL
Vitamine 100x	10 mL
Spurenelemente 500x	2 mL
deion. H ₂ O	568 mL
Aseptisch vereinigt.	

In Medien, in denen die Succinatkonzentration variiert wurde, wurde das Verhältnis Succinat 10x zu deion. H₂O entsprechend angepasst.

2.2.3.3 Anzucht aerober Vorkulturen

Für alle wachstumsrelevanten Experimente mit *D. shibae* wurden zunächst mit einer sterilen Impföse Bakterien aus einer Glycerinkultur entnommen und auf einer MB-Agar-Platte im Dreistrich-Verfahren ausgestrichen. Die Platte wurde für zwei Tage bei 30 °C und im Anschluss bis zu sieben Tage bei Raumtemperatur dunkel inkubiert. Anschließend wurden Bakterien mit einer sterilen Impföse aus einer einzelnen Kolonie entnommen und in ein steriles Reagenzglas mit 5 mL SWM überführt. Dieses wurde für 30 h bei 30 °C und 200 rpm dunkel inkubiert. Letztendlich wurden aus dieser Kultur 0,6 mL in einen sterilen 50 mL-Schikanekolben mit 30 mL SWM überführt und dieses wieder bei 30 °C und 200 rpm dunkel inkubiert. Nach dieser Prozedur erhielt man stets eine sich in der exponentiellen Wachstumsphase befindliche Kultur mit einer optischen Dichte von etwa $e_{650} = 1,6$, die für weitere Untersuchungen verwendet werden konnte.

2.2.3.4 Kultivierung in Mikrotiterplatten

Für Mikro-Wachstumsexperimente wurden Mikrotiterplatten mit 24 Kavitäten (Nunc, Katalog-Nummer 144530) verwendet. Die einzel-

nen Kavitäten wurden mit 1160 μL SWM befüllt und mit 40 μL einer frischen Vorkultur inokuliert. Wenn nicht anders angegeben, hatte das SWM eine Endkonzentration von 2 g/L Succinat. Um Referenzmessungen vorzunehmen, wurde immer je Replikatreihe eine Kavität mit 1160 μL SWM und, anstatt es zu inokulieren, mit 40 μL deion. H_2O gefüllt. Anschließend wurde die Mikrotiterplatte mit einer luftdurchlässigen Folie (BreathEasy, Roth, Artikelnummer T093.1) abgedeckt und verklebt. Die Inkubation erfolgte bei 30 °C mit 480 rpm und einem orbitalen Schüttelradius von 6 mm in dem Multi-Label Lesegerät VarioskanFlash.

Während dieser Kultivierung wurde das Wachstum in regelmäßigen Abständen durch Messung der optischen Dichte im Lesegerät bestimmt. Hierzu wurde in den jeweiligen Kavitäten die Extinktion bei 650 nm ermittelt und die Extinktion der zugehörigen Referenzkavität abgezogen. Da während Kultivierungen über lange Zeiträume von einer Verdunstung des Mediums und entsprechend einer Aufkonzentrierung der Zellen in der Kultur auszugehen war, wurde bei jeder Messung die Schichtdicke der Proben ermittelt, um zeitlich vergleichbare Extinktionsswerte zu erhalten. Hierzu wurde in den Referenzkavitäten die Extinktion e bei 900 nm und 977 nm gemessen und die Schichtdicken h entsprechend der Formel

$$h = \frac{e_{977} - e_{900}}{k} \quad (2.8)$$

berechnet. Hierbei ist k ($[k] = 1/\text{cm}$) eine von Medium, Folie und Mikrotiterplatte abhängige Konstante, die sich durch eine Kalibrierung über verschiedene Schichtdicken ermitteln lässt (Held, 2009). Da sich das Verfahren zur Schichtdickenbestimmung ausschließlich für wässrigen Proben eignet, also solchen die keine streulichtbedingte Extinktion aufweisen, wurde die Annahme getroffen, dass die Referenzkavitäten keine streulichtbedingte Extinktion aufweisen.

renzkavitäten die gleiche Verdunstung wie die Proben der zugehörigen Replikatreihe erfahren. Letztendlich ergab sich die auf eine Schichtdicke von 1 cm bezogene optische Dichte einer Probe r mit zugehöriger Referenz b aus

$$e_{650\text{nm}}(r) = \frac{e_{650}(r) - e_{650}(b)}{h(b)} \quad (2.9)$$

2.2.3.5 Lichtbedingungen während der Mikrotiterplattenkultivierung

Um während der Kultivierung in Mikrotiterplatten den Umweltparameter *Helligkeit* zu variieren, wurde eine funkgesteuerte Elektronik in eine Mikrotiterplatte integriert (Abbildung 2.1). Die hierzu benötigten Bauteile sind Tabelle 2.3 zu entnehmen.

Es war notwendig die Kavitäten A5, A6, B5, B6 und C1 – D6 samt dem Boden der Platte herauszusägen. Ein neuer passgenauer Boden aus Polyvinylchlorid wurde anschließend eingesetzt und mit Silikondichtmasse an der Platte verklebt. Als Leuchtmittel kam ein Subminiaturlampe zum Einsatz. Diese wurde auf einem Kunststoffsteg so angebracht, dass sie die Kavitäten B3 und B4 vollkommen gleich beleuchtete. Als Schalter für das Leuchtmittel kam eine Funksteuerungsplatine zum Einsatz, die per externem Sender das Signal zum Ein- und Ausschalten erhielt. Die Stromversorgung wurde durch einen Lithium-Polymer-Akkumulator bereitgestellt, der in Parallelschaltung sowohl den Funkempfänger wie auch das Leuchtmittel versorgte (Abbildung 2.2). Die Kavitäten A1 – A4 waren durch ein 67 mm×17 mm großes Stück Aluminiumfolie abgeschattet zuzulassen (Abbildung 2.1).

Gesteuert wurde der Funkempfänger über den USB- (*engl.*: Universal Serial Bus) Funksender FS20 PCS, der sich durch das HID- (*engl.*: Human Interface Device) Protokoll steuern lässt. Zum Ein-

Tabelle 2.3: Bauteile zur Beleuchtung einer Mikrotiterplatte.

Bauteil	Merkmal	Artikelnummer / Hersteller
Mikrotiterplatte	24 Kavitäten, Polystyren, unbehandelt	144530 Nunc, Langensbold, Deutschland
Funkempfänger	FS20, 868,35 MHz	84084 ELV Elektronik AG Leer, Deutschland
Funksender	FS20-PCS, 868,35 MHz, 100 m Reichweite, USB-HID	85861 ELV Elektronik AG Leer, Deutschland
Leuchtmittel	Subminiaturlampe, 12 V, 30 mA, 3,2 mm×6,5 mm	21109123 Barthelme GmbH Nürnberg, Deutschland
Akkumulator	Lithium-Polymer, 3 Zellen, 11.1 V, 450 mAh, (55×29×15,5) mm	KOK4503S1P Kokam Co., Ltd, Kyunggi-Do, Südkorea

schalten des Lichts wurde das 11 byte lange HID-Signal 00 06 F1 00 10 0F 10 00 00 00 00 und zum Ausschalten das Signal 00 06 F1 00 10 0F 00 00 00 00 00 an den Empfänger gesendet. Zuvor wurde der Empfänger einmalig mit dem Hauscode 00 10 und der Geräteadresse 0F initialisiert.

2.2.3.6 Synchrone Steuerung der Mikrotiterplattenbeleuchtung

Während einer Kultivierung im Multi-Label Lesegerät wurden, wenn nicht anders angegeben, die einzelnen Kavitäten alle 30 min photometrisch vermessen. Um eine zuverlässige Messung zu gewährleisten, war es notwendig, während der Kultivierung für strikte, dunkle Be-

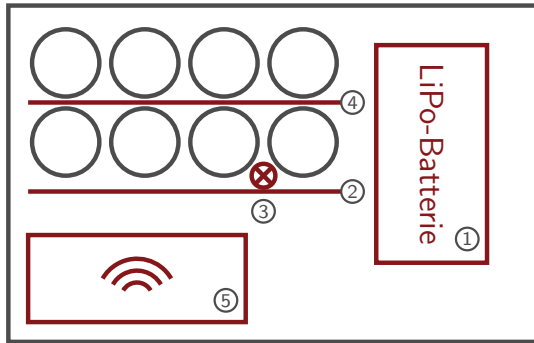


Abbildung 2.1: Schematische Darstellung der umgebauten Mikrotiterplatte mit integrierter Belichtungselektronik. Schematisch grau dargestellt sind die Reste der originalen Mikrotiterplatte, rot dargestellt sind die integrierten Erweiterungen: (1) Lithium-Polymer Akkumulator, (2) Steg zur Befestigung des Leuchtmittels, (3) Subminiaturlampe, (4) Aluminiumfolien-Abschattung, (5) FS20 Funkempfänger.

dingungen in der Messkammer zu sorgen. In Versuchen bzw. Phasen, in denen die Mikrotiterplatte beleuchtet wurde, musste das Leuchtmittel daher während der Messung ausgeschaltet werden. Dies wurde über eine synchronisierende Steuerung des Funkschalters und der einzelnen Programmaufrufe gewährleistet, indem diese Aufrufe über die Automatisierungsschnittstelle des Lesegerätes erfolgten.

2.2.4 Bacteriochlorophyll Visualisierung und Quantifizierung

2.2.4.1 Licht- und Fluoreszenzmikroskopische Untersuchungen

Zur lichtmikroskopischen Untersuchung von *D. shibae* wurden 5 μL einer Flüssigkultur auf einen Objektträger gegeben und mit einem Deckglas abgedeckt. Die Probe wurde anschließend unter Zuhilfe-

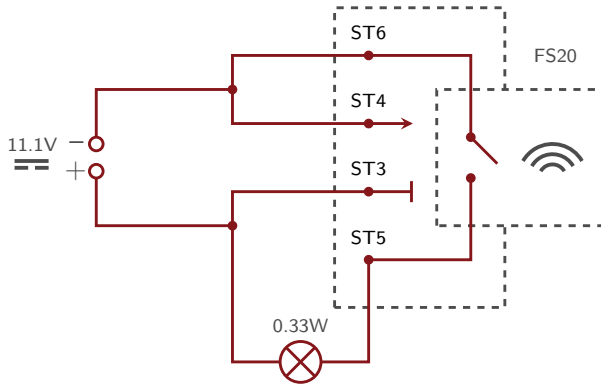


Abbildung 2.2: Schaltkreis der funkgesteuerten Elektronik zur Beleuchtung der Mikrotiterplatte. Der Funkschalter ist mit der grau gestrichelten Linie angedeutet und mit FS20 bezeichnet. Die Bezeichnung der Anschlüsse ST3 - ST6 entspricht der der Lötstellen auf der Platine. In den Schaltkreis ist eine Miniaturglühlampe mit einer Nennleistung von 360 mW integriert. Ein Lithium-Polymer-Akkumulator mit einer Ausgangsspannung von 11.1 V stellt die Stromversorgung bereit. Durch die von der Verbrauchsspannung der Subminiaturlampe abweichende Ausgangsspannung der Stromquelle reduziert sich die Leistung des Leuchtmittels am Betriebspunkt auf 330 mW.

nahme von Immersionsöl und einem 100x-Objektiv im Lichtmikroskop vergrößert. Mit Hilfe der Software¹⁰ AxioVision 4.8 fand eine Visualisierung und Speicherung der Aufnahmen statt.

Für fluoreszenzmikroskopische Aufnahmen wurde unter gleichen Bedingungen ein geeigneter Reflektorrevolver in den Strahlengang zwischen Objektiv und Okular bzw. Lichtquelle gebracht. Visualisiert wurde die Fluoreszenzemission des BChl-*a* bei über 850 nm nach Exzitation an der ersten Q_y -Bande bei 805 nm (Abbildung 2.3). Die Belichtungszeit betrug hierbei 4 s. Die Aufnahmen erfolgten ebenfalls durch die Software AxioVision.

¹⁰<http://www.zeiss.de/axiovision>

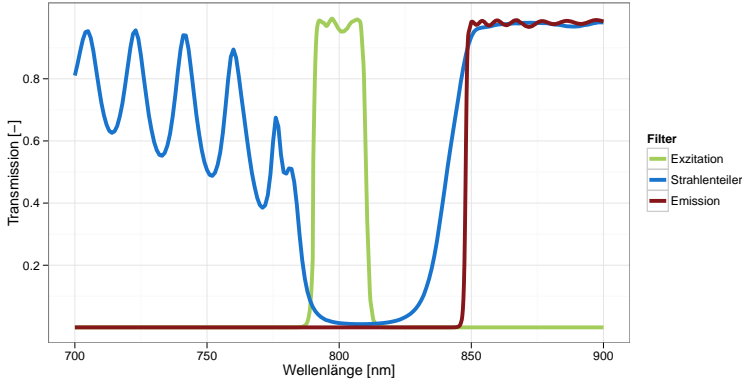


Abbildung 2.3: Transmissionsspektren (AHF Analysetechnik AG, Tübingen) des verwendeten Reflektormoduls zur BChl-*a* - Visualisierung im Fluoreszenzmikroskop. Das Spektrum des Exitationsfilters ist grün, das des Strahlenteilers blau und das des Emissionsfilters rot eingezeichnet.

2.2.4.2 *In vivo*-Quantifizierung von Bacteriochlorophyll

Zur *in vivo*-Quantifizierung des BChl-*a* wurden Proben aus einer planktonisch wachsenden Kultur von *D. shibae* entnommen, und die Extinktion der Q_y -Bande e_{870} bei 870 nm gemessen. Da die Messung direkt in der trüben Kulturflüssigkeit stattfand, musste die Extinktion der Partikelstreuung vom jeweiligen Messwert Wert abgezogen werden, um nur den Absorptionsanteil a_{870} des BChl-*a* zu ermitteln. Hierzu wurde die Extinktion der Partikelstreuung s_{870} an der Wellenlänge 870 nm durch eine lineare Basislinienberechnung zwischen den Wellenlängen 770 nm und 900 nm gemäß der Formel

$$a_{870} = e_{870} - \frac{e_{900} + 0,3 \cdot e_{770}}{1,3} \quad (2.10)$$

interpoliert (Kapitel 3.2.2.2). Die so ermittelten Werte wurden auf die gemessene Extinktion bei 650 nm bezogen, um den BChl-*a* Gehalt pro Zelldichte semi-quantitativ zu bestimmen.

3 Ergebnisse und Diskussion

3.1 eSOMet – Eine neue Software zur Metabolomdatenanalyse

3.1.1 Ein verbessertes Verfahren zur Metabolomdatenauswertung

Zur Auswertung komplexer Metabolomexperimente und deren multivariater Daten wurde das Programm eSOMet entwickelt (Haddad *et al.*, 2009). In ihm wurden typische multi- und univariate Methoden implementiert und zu einer neuartigen Auswertestrategie verknüpft. Das Ziel war es die Datenexploration auf vereinfachte und intuitive Weise vor einem starken statistischen Hintergrund zu ermöglichen. Zudem wurde die Methode der emergenten selbstorganisierenden Karten implementiert, deren Mehrwert in der Metabolomik bisher ununtersucht war.

Das Ziel war es ein Werkzeug zur Verfügung zu stellen, welches die metabolische Profile organisieren und diese clustern kann. Zum Clustern wird optional die Möglichkeit geboten, die metabolischen Profile mittels Hauptkomponentenanalyse zuvor auf einen dimensionsreduzierten Raum zu projizieren. Dieses Vorgehen reduziert die Komplexität der Daten und beschleunigt dadurch die Laufzeit der nachgeschalteten Clusteralgorithmen. Ein ähnliches Verfahren wurde in der Vergangenheit bereits vorgeschlagen: Kouskoumvekaki

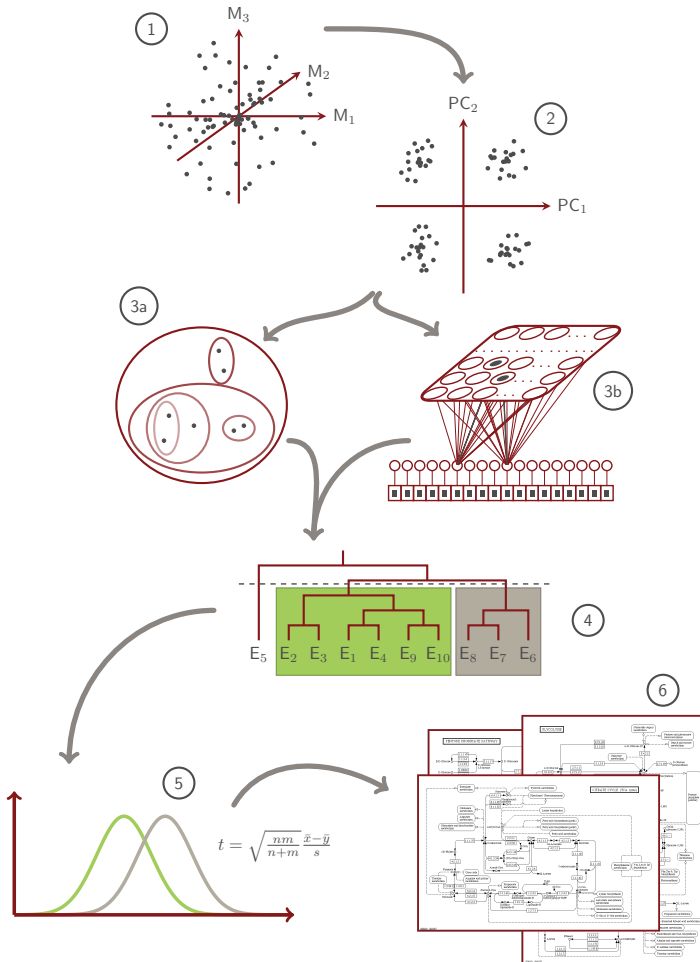


Abbildung 3.1: Ein typischer Arbeitsfluss der Metabolomdatenauswertung in eSOMet. (1) Laden der multivariaten Daten und ggf. Normalisierung. (2) Projektion der Daten auf einen dimensionsreduzierten PCA-Raum. (3) Clustern der metabolischen Profile mittels HCA (3a) oder ESOM-Analyse (3b). (4) Manuelle Auswahl geeigneter Cluster aus dem erzeugten Dendrogramm. (5) Berechnung der statistischen Signifikanz der Varianzen einzelner Metabolitabundanz zwischen den Clustern. (6) Visualisierung der Ergebnisse auf metabolischen Netzwerkarten.

et al. (2008) z.B. entfernen aus den metabolischen Profilen alle Metabolite, die nach einer Varianzanalyse global nicht signifikant variiert sind und führen das Clustern lediglich auf Basis der verbleibenden Metabolite durch. Auch wenn diese Methode einem rationalen Ansatz folgt, birgt sie problematische Aspekte. Sie berücksichtigt zum einen nicht, dass Ausreißer die Varianzanalyse enorm stören können. Zum anderen birgt sie die Gefahr, dass das Clustern ausschließlich auf Basis von Unähnlichkeit erfolgt und die Ähnlichkeiten zwischen den Profilen keinen Einfluss haben. Die Reduktion der Daten per PCA hingegen, erhält deren inherente Struktur und entfernt lediglich den Beitrag jedes Metaboliten zum globalen, uncharakteristischen Rauschen.

Im weiteren kommt dem Clustern eine entscheidende Rolle innerhalb der Auswertestrategie zu Teil. Sie gibt dem Anwender nicht nur die Möglichkeit die Ähnlichkeiten zwischen den metabolische Profilen zu bestimmen, sondern lässt gleichzeitig auch die einfache Identifikation von Ausreißern zu, um diese von weiteren Analysen auszuschließen. Weiterhin ermöglicht die Software die manuelle Auswahl einzelner Cluster und die Bestimmung statistisch signifikant variiert Metabolitabundanzen zwischen den Clustern.

Letztendlich integriert eSOMet den jeweils aktuellen Datenbestand der KEGG PATHWAYS (Kanehisa & Goto, 2000) und KEGG LIGAND (Goto *et al.*, 2002) Datenbanken. Damit ist es möglich, die Ergebnisse der statistischen Auswertung auf den hinterlegten Reaktionsnetzwerken abzubilden und somit eine visuelle Beurteilung der Experimente in einem biologischen Kontext durchzuführen. Ein typischer Arbeitsfluss, wie er mit der Software vorgenommen werden kann, ist in Abbildung 3.1 dargestellt.

Grundsätzlich existiert bereits eine Reihe bioinformatischer Werkzeuge, die statistische Methoden mit der Visualisierung auf bioche-

mischen Netzwerken vereinen. Für Genexpressionsprofile eignen sich bspw. PathMAPA (Pan *et al.*, 2003), PathwayExplorer (Mlecnik *et al.*, 2005) oder Pathway Processor (Grosu *et al.*, 2002); für metabolische Profile MetNet (Wurtele *et al.*, 2003), KaPPA-View (Toki-matsu *et al.*, 2005), MAPMAN (Thimm *et al.*, 2004) oder VANTED (Junker *et al.*, 2006). Diese Programme haben jedoch alle gemein, dass sie bereits intensiv prozessierte Daten erfordern. Weiterhin beschränkt sich die statistische Auswertung auf Varianzanalysen zur Detektion signifikant variierten Transkripte bzw. Metabolite. Zumindest VANTED unterstützt auch das Clustern von Metaboliten nach ähnlichen differentiellen Änderungen über die experimentellen Faktoren. Keines dieser Programme erlaubt jedoch das Clustern metabolischer Profile, welches im Zentrum des eSOMet-Prozesses steht.

3.1.2 Das Softwaredesign

eSOMet ist ein rein Java basiertes Softwareprodukt und liefert neben der Implementierung der statistischen und mathematischen Prozeduren eine grafische, leicht zu bedienende und intuitive Benutzeroberfläche. Die globale Architektur der objektorientierten Programmierung richtete sich strikt nach dem Model-View-Controller-Prinzip (Gamma *et al.*, 2001). Das heißt, dass Anwendungslogik, Präsentationsschicht und Steuerung voneinander getrennt implementierte Teile sind, deren Kommunikation untereinander über definierte Schnittstellen stattfindet. Dies macht das Softwarepaket flexibel für Erweiterungen Dritter und lässt ebenso zu, nur die Anwendungslogik als Programmierbibliothek in weiteren Softwareprodukten zu nutzen.

Teile von eSOMet, nämlich die Anwendungslogik zur Berechnung der ESOMs, dem Zeichnen von Dendrogrammen, der graphischen Darstellung von Metabolitveränderungen und die Abbildung dieser

auf den KEGG Netzwerkkarten wurden durch frühere Arbeiten (Hiller, 2006) zur Verfügung gestellt, erforderten aber weiterhin Anpassungen, Laufzeitoptimierungen und Fehlerbereinigungen.

Die Benutzeroberfläche (Abbildung 3.2) stellt alle wesentlichen Funktionen der Analysestrategie (Abbildung 3.1) zur Verfügung. Sie gliedert sich zum einen in die Anzeige der Projektstruktur, die alle Objekte der aktuellen Auswertung, wie z.B. Rohdaten, PCA, Ergebnisse des Clusters usw. übersichtlich darstellt. Zum anderen wird in einem weiteren Anzeigenbereich das jeweils ausgewählte Objekt in einer geeigneten Präsentation dargestellt. So ist es möglich schnell zwischen den Zwischenergebnissen zu wechseln und Vergleiche zwischen Auswertungen mit verschiedenen Parametern durchzuführen. Ferner wurden für die komplexen Analysen, die eine Vielzahl von Parametern erfordern, leicht auszuführende Programmassistenten entwickelt, die durch die verschiedenen Einstellungen leiten und stets sinnvoll voreingestellte Parameter anbieten (Abbildung 3.3).

Um ein Projekt in eSOMet zu starten, muss dieses zunächst angelegt werden und alsdann die multivariate Datenmatrix geladen werden. Hierzu besteht sowohl die Möglichkeit eine Tabelle aus einem Tabellenkalkulationsprogramm per „Kopieren & Einfügen“ zu importieren oder eine Datei im CSV- (*engl.*: Character Separated Values) Format zu laden. Ist im Anschluss an die multivariate Auswertung eine Visualisierung der Daten auf den KEGG Netzwerkkarten gewünscht, ist es erforderlich in den importierten Daten die zu jedem Metabolit gehörende KEGG-ID anzugeben. Dieser Bezeichner ist innerhalb der gesamten KEGG LIGAND Datenbank voll qualifizierend und erlaubt eine eindeutige Zuordnung. Das Projekt kann zu jedem Zeitpunkt mit allen ausgeführten Analysen in einem binären Format gespeichert werden, um es zu einem späteren Zeitpunkt wieder zu laden.

The screenshot shows the eSOMet software interface. On the left, a tree view displays the project structure under 'neu', including 'Data' (with 'Original data' selected), 'PCA', 'ESOM', and 'HCA'. On the right, a table titled 'Original data' displays metabolite data. The table has columns for Metabolite, KeggID, and various scaled data columns (Glu-0h-a to Glu-2h-a).

Metabolite	KeggID	Glu-0h-a	Glu-0h-b	Glu-0h-c	Glu-1h-a	Glu-1h-b	Glu-1h-c	Glu-2h-a
2-Aminob...	C00334	817E001	664E001	665E001	992E001	104E002	944E001	755E001
2-Deoxy...	C00673	000E000	241E000	144E000	215E000	143E000	166E000	809E-001
2-Hydrox...	C01087/...	315E003	293E003	324E003	848E002	879E002	782E002	405E002
2-Hydrox...	C02502	194E002	273E002	296E002	258E002	318E002	193E002	203E002
2-Oxogluc...	C00026	653E003	621E003	791E003	156E003	156E003	158E003	845E001
3-Phosph...	C00631	000E000	000E000	000E000	000E000	000E000	000E000	000E000
3-Phosph...	C00197	486E002	965E002	123E003	816E002	107E003	142E003	313E002
5-Amino...	C00430	000E000	296E000	246E000	164E000	143E000	198E000	330E-001
5-Oxopro...	C01879	703E002	700E002	678E002	661E002	703E002	589E002	511E002
6-Phosph...	C00345	173E000	121E000	198E000	258E000	431E000	651E000	454E000
Adenine	C00147	839E002	508E002	473E002	359E002	248E002	853E002	758E001
Alanine	C01401/...	159E003	141E003	157E003	105E003	116E003	102E003	878E002
alpha_alp...	C01083	231E004	204E004	199E004	270E004	338E004	227E004	169E004
AMP	C00020	000E000	000E000	000E000	000E000	000E000	000E000	000E000
Arginine	C02385	000E000	000E000	000E000	000E000	000E000	000E000	000E000
beta-Alan...	C00099	119E002	165E002	186E002	240E001	224E001	192E001	587E001
C15acid	-	319E002	243E002	201E002	208E002	423E002	207E002	166E002
C16acid	-	319E003	242E003	200E003	275E003	374E003	304E003	381E003
C31acid	-	000E000	000E000	000E000	000E000	000E000	000E000	000E000
C9acid	-	404E001	543E001	388E001	591E001	435E001	443E001	600E001
cis9octad...	-	000E000	000E000	000E000	000E000	000E000	000E000	000E000
Citrate	C00158	278E001	412E001	430E001	867E001	161E002	126E002	190E002
Cyclohexa...	C00854	469E001	811E001	757E001	926E001	124E002	485E001	989E001
Cyclolucid...	-	345E002	162E000	385E-001	336E000	276E000	248E000	519E000
Cystathion...	C00542	000E000	000E000	000E000	000E000	000E000	000E000	000E000
Cytosine	C00380	998E001	964E001	969E001	905E001	912E001	861E001	614E001
D-Fructos...	C00354	753E-001	951E000	415E000	253E000	134E001	118E001	406E000
D-Fructos...	C00085	314E002	456E002	509E002	544E002	451E002	695E002	378E002

Abbildung 3.2: Bildschirmfoto der Programmoberfläche von eSOMet. Der linke Bereich stellt die Projektstruktur der aktuellen Auswertung dar und beinhaltet alle zwischengespeicherten Ergebnisse. Der rechte Bereich stellt die Objekte, die links ausgewählt werden können, in einer jeweils geeigneten Form dar. So sind in diesem Bildschirmfoto rechts die Rohdaten des aktuell geladenen Metabolomdatendatei tabellarisch dargestellt.

Um die Visualisierung auf den KEGG Netzwerkarten vorzunehmen ist es nötig, dass der Benutzer die relevanten Daten der KEGG LIGAND und KEGG PATHWAYS Datenbanken in das Projekt lädt. Die Daten wurden aus zwei Gründen bewusst nicht fest in das Programm integriert:

1. Verzeichnete die KEGG Datenbank schon immer eine hohe Dynamik bezüglich ihres Datenbestandes und wird ständig um weitere Informationen erweitert. Eine feste Integration barg die Gefahr, dass die Aktualität des Datenbestandes in eSOMet

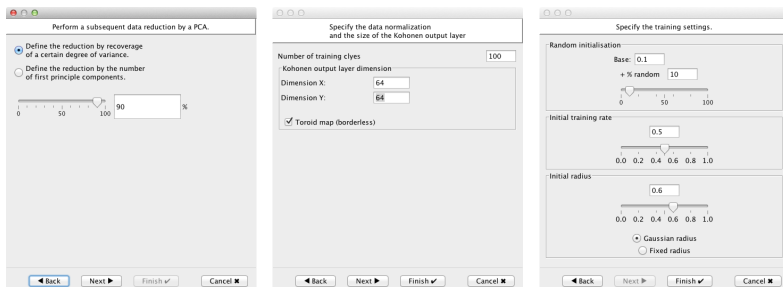


Abbildung 3.3: Darstellung dreier Programmfenster des Programmassistenten zur Berechnung der ESOMs. Links: Optionen der Datenreduktion durch eine PCA. Mitte: Definition der Ausgabeschicht der ESOM. Rechts: Parameter des ESOM Trainingszyklus.

von seinem Aktualisierungszyklus abhängt.

2. Desweiteren steht die Nutzung der KEGG Daten unter einer nicht freien Lizenz. Eine Integration der Daten hätte daher die kostenpflichtige Lizenzierung der Daten erfordert.

Daher bietet eSOMet die Möglichkeit per Programmassistenten die Daten als ASCII-Dateien vom KEGG FTP-Server zu laden, nachdem der Benutzer einen individuellen Lizenzvertrag¹ mit den Lizenzgebern abgeschlossen hat. Desweiteren erzeugt der Programmassistent eine interne Datenbank und befüllt diese mit den heruntergeladenen Daten. Die Datenbank steht von da an dem Programm zur Verfügung, um den Zusammenhang zwischen den metabolischen Daten und den hinterlegten Informationen herzustellen. Die Umsetzung als SQL Datenbank bietet zudem optimale Laufzeiteffizienz und ermöglicht so ein flüssiges Arbeiten.

eSOMet ist unter der Lesser General Public License 3.0 lizenziert.

¹<http://www.bioinformatics.jp/en/keggftp.html>

Das Programm und sein gesamter Quellcode können von der Projektwebseite² heruntergeladen werden. Dort finden sich auch ein detailliertes Handbuch und ein Testdatensatz auf dessen Auswertung im Folgenden weiter eingegangen wird.

3.1.3 Evaluierung von eSOMet mit einem Metabolomdatensatz von *Corynebacterium glutamicum* Kultivierungen

eSOMet wurde anhand eines Metabolomdatensatzes evaluiert, der im Rahmen einer weiteren Arbeit erzeugt (Frimmersdorf, 2010) und freundlicherweise zur Verfügung gestellt wurde. Die metabolischen Profile stammten aus drei Fermenterkultivierungen von *Corynebacterium glutamicum*, die jeweils mit Glucose (20 g/L), Fructose (20 g/L) und Actetat (16 g/L) als einziger Kohlenstoffquelle durchgeführt worden sind. Aus den Kulturen wurden nach Inokulation stündlich über einen Zeitraum von bis zu 12 h - 14 h Proben in Triplikaten entnommen und per GC-MS-Analyse vermessen. Insgesamt setzt sich der Datensatz somit aus 126 Profilen zusammen, in denen insgesamt 209 distinkte Metabolite chemoinformatisch identifiziert werden konnten. Die gemessenen Metabolitabundanzen waren semi-quantitativ (Kapitel 1.2.1). Zudem wurde zu jedem Probenahmezeitpunkt die Zelldichte photometrisch bestimmt, um einen Rückschluss auf die jeweilige Wachstumsphase ziehen zu können (Abbildung 3.4).

Der Datensatz eignet sich aufgrund seines faktoriellen Designs hervorragend zur Evaluierung der Software im Allgemeinen und der verschiedenen Clustermethoden im Speziellen. So ist bei dem Clustern der metabolischen Profile sowohl eine Differenzierung nach verwerteter Kohlenstoffquelle, als auch nach der jeweiligen Wachstumsphase

²<http://esomet.tu-bs.de/>

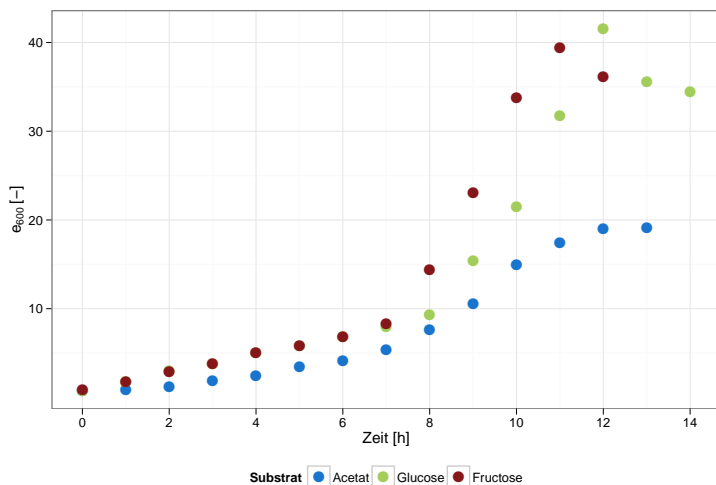


Abbildung 3.4: Wachstumskurven der *C. glutamicum* Kultivierungen. Die Zelldichte wurde über die optische Dichte des Kulturmediums bei 600 nm stündlich gemessen. Blau ist die Kultivierung mit Acetat, grün mit Glucose, und rot mit Fructose als einziger Kohlenstoffquelle dargestellt (Frimmersdorf, 2010).

zu erwarten. Welcher Einfluss letztendlich größer ist und welche Metabolite auf die jeweiligen Diskriminierungen den größten Einfluss haben, wird im Folgenden beschrieben.

PCA Der Erste Schritt, der sich nach dem Laden der Metabolomdaten anbietet, ist eine visuelle Kontrolle der PCA-projizierten Daten (Abbildung 3.5 a). Für den vorliegenden Datensatz führte die Projektion der Profile auf die ersten beiden Hauptkomponenten zu einer strikten Separation nach der verwendeten Kohlenstoffquelle. Eine solche frühe Beobachtung ist bspw. ein guter Indikator dafür, dass die globale Varianz der verschiedenen Stoffwechselmodi

eher durch die zur Verfügung gestellte Kohlenstoffquelle, als die Wachstumsphase beeinflusst wird. Desweiteren liefert die Anzeige der durch die Hauptkomponenten kumulativ erklärten Varianzen einen guten Hinweis auf die Qualität der Daten (Abbildung 3.5 b). Ein steiler Anstieg deutet hierbei an, dass viele Metabolite untereinander den gleichen differenziellen Änderungen unterliegen. Da dies grundsätzlich ein Umstand ist, der bei dem vorliegenden experimentellen Design zu erwarten ist, würde ein flacher Anstieg hingegen auf eine hohe nicht induzierte bzw. technische Varianz hindeuten. Hier ließ sich jedoch 90% der Varianz bereits durch die ersten 19 von insgesamt 204 Hauptkomponenten erklären, was auf einen sehr hohen Beitrag biologisch induzierter Varianz hindeutete.

ESOM-Analyse Im nächsten Schritt wurden die Profile durch ESOMs geclustert. Die Ausgabeschicht wurde hierbei 64×64 Knoten groß gewählt. Das Training der ESOM erfolgte über $t_{\max} = 100$ Trainingszyklen mit einer initialen Trainingsrate von $\alpha = 0,5$ und einem initialen Gauss'schen Trainingsradius von $\nu = 0,6$ (Kapitel 2.1.3.5). Das Training erfolgt mit den auf die ersten 19 Hauptkomponenten projizierten Daten. Nach der Abgeschlossenen Analyse zeigt eSOMet zwei neue Einträge im Projektbaum an: Die U^* -Matrix der trainierten ESOM mit den eingebetteten Profilen und die Transformation dieser zu einem Dendrogramm.

Bereits an der U^* -Matrix (Abbildung 3.6) lässt sich wiederholt erkennen, dass die metabolischen Profile in erster Linie nach der zur Verfügung gestellten Kohlenstoffquelle geclustert worden sind. Das feinkörnige Clustern lässt sich dem Dendrogramm entnehmen, welches aus der U^* -Matrix überführt wurde (Anhang 4.2). Aus dem Dendrogramm können eine Reihe von Aspekten, die dem Metabolomdatensatz zugrundelagen entnommen werden:

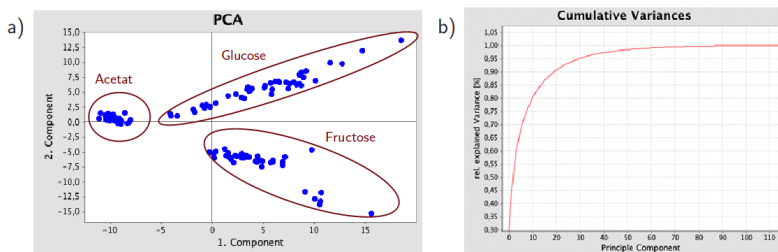


Abbildung 3.5: Ausschnitt eines Bildschirmfotos der PCA-Oberfläche in eSOMet. **a)** Projektion der metabolischen Profile auf die ersten beiden Hauptkomponenten PCA. Die Profilbezeichner wurden in eSOMet der Übersichtlichkeit halber ausgeblendet und die Zugehörigkeit der Profile nachträglich eingezeichnet (Rot). Offensichtlich trennen erste und zweite Hauptkomponente die metabolischen Profile strikt nach der jeweiligen Kultivierung mit den drei verschiedenen Kohlenstoffquellen Glucose, Fructose und Acetat. **b)** Anzeige der kumulativ erklärten Varianz durch die ersten Hauptkomponenten. In dem vorliegenden Datensatz lässt sich bereits 90% der globalen Varianz durch die ersten 19 Hauptkomponenten erklären.

1. Alle Profile der Acetat-Kultivierung bilden ein exklusives Cluster, das beinahe auf höchster Ebene des Dendrogrammes gebildet wird. Die Profile der Glucose- und Fructose-Kultivierung bilden parallel dazu ein eigenes Cluster. Dieses Ergebnis deckt sich mit bekanntem Wissen über den Stoffwechsel von *C. glutamicum*: Glucose- und Fructose Metabolismus unterscheiden sich lediglich in der Aufnahme der Zucker und den initialen Stoffwechselschritten. Der eigentliche Hexose-Katabolismus durch die Glycolyse ist allerdings bei Wachstum auf beiden Substraten identisch. Die aktivierten Bereiche des metabolischen Netzwerkes unterscheiden sich im Gegensatz dazu im Acetat-

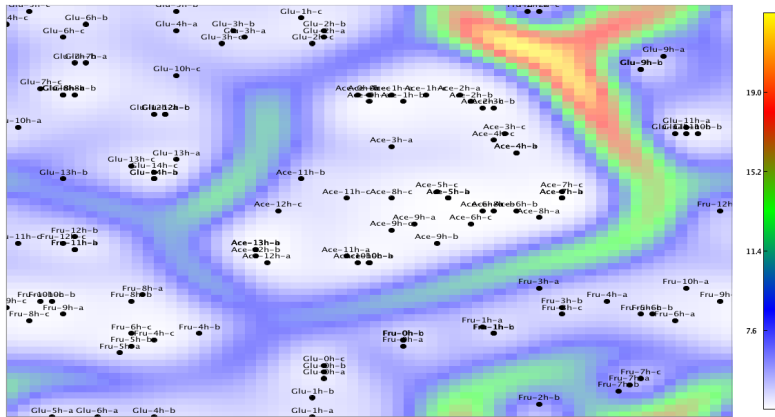


Abbildung 3.6: Ausschnitt eines Bildschirmfotos der U*-Matrix Oberfläche in eSOMet. Dargestellt ist die U*-Matrix einer trainierten ESOM. Für jeden der 64×64 Knoten der Ausgabeschicht wurde die euklidische Distanz zu seinen vier unmittelbaren Nachbarn berechnet. Die räumliche Verteilung dieser Distanzen erzeugt eine drei-dimensionale Landschaft, in die metabolischen Profile nach abgeschlossenen Training der ESOM eingebettet sind. Zur besseren Darstellung wurde eine zwei-dimensionale Darstellung gewählt, wobei die Distanzen durch eine eindeutige Farbskala (rechts) beschrieben werden. Hierbei diskriminieren gelb bis rote Bereiche Profile mit einer hohen Dissimilarität, während sich besonders ähnliche Profile in den gleichen *Tälern*, also den zusammenhängenden blau bis weißen Bereichen, gruppieren. Die Bezeichner entsprechen der Syntax <Substrat> – <Probenahmezeitpunkt> – <Replikat>. „Ace“ steht hierbei für Acetat, „Glu“ für Glucose und „Fru“ für Fructose.

Katabolismus deutlich (Wendisch *et al.*, 2000).

2. Entsprechend separieren sich auch im Glucose-Fructose-Cluster die Profile nach dem verwendeten Substrat relativ strikt. Ausnahmen bilden die Profile der Proben Fru-2h-a - c, Fru-7h-a - c, Fru-3h-a, Glu-9h-a - c und Glu-13h-b (die Syntax der Bezeichner ist in Abbildung 3.6 erklärt). Sie finden sich auf verhältnismäßig hohen Hierarchieebenen des Dendrogramms wieder und ließen sich somit nicht eindeutig zuordnen.
3. In den meisten Fällen finden sich die Proben, die eine technische Replikatgruppe bilden, in einem Cluster auf unterster oder zumindest niedriger Hierarchieebene wieder. Das bestätigt die grundsätzliche Annahme, dass statistisches Rauschen stets den geringsten Einfluss auf die Varianz der Proben hat.
4. Letztendlich lässt sich über das gesamte Dendrogramm auch die Tendenz erkennen, dass Proben gemäß der jeweiligen Wachstumsphase zueinander geclustert bzw. voneinander diskriminiert wurden. Dies im einzelnen qualitativ auszuwerten ist bei der Komplexität der Daten jedoch wenig sinnvoll. Daher ist auf Kapitel 3.1.4 verwiesen, wo das hierarchische Clustern mittels einer mathematischen Methode beurteilt wird.

Differentielle Änderungen der Metabolitabundanzen Im letzten Schritt des Analyseflusses wurden die differentiellen Änderungen der Metabolitabundanzen von der Glucose- zur Fructose-Kultivierung ausgewertet. Hierzu konnten im Dendrogramm die jeweiligen größten Cluster ausgewählt werden (Anhang A1). Exemplarisch soll diese Auswertung anhand des Pentosephosphat Metabolismus detailliert betrachtet werden (Abbildung 3.7). Die signifikanten Änderungen, die in den gemessenen Abundanzen ersichtlich sind, lassen sich



mehr als doppelt so hoch, als in der Fructose Kultivierung. Desweiteren führt der Mangel an Fructose-1,6-bisphosphatase Aktivität zu reduzierten Flüssen durch das Pentosephosphat Netzwerk, weshalb die Abundanzen von Xylulose-5-phosphat, Ribose-5-phosphat und Ribulose-5-phosphat in der Glucose-Kultivierung signifikant zwei bis drei mal höher gewesen sind.

Durch die Reproduktion dieser Ergebnisse konnte der Mehrwert und die Zuverlässigkeit von eSOMet unter Beweis gestellt werden.

3.1.4 Quantitativer Vergleich der verschiedenen Clusterverfahren

Zuletzt sollte noch untersucht werden, inwiefern die Methode der ESOMs gegenüber der klassischen HCA im Bereich der Metabolomik einen Mehrwert bringt. Hierzu wurde der gleiche Datensatz in eSOMet zusätzlich einer HCA unterzogen. Um die Bedingungen im Vergleich zur ESOM-Analyse so vergleichbar wie möglich zu gestalten, wurde auch hier der Datensatz zunächst auf die ersten 19 Hauptkomponenten der PCA reduziert. Als Distanzfunktion wurde die Methode *Durchschnittlich Entfernter Nachbar* und als Metrik die Euklidische Distanz gewählt (Kapitel 2.1.3.4).

An dem erzeugten HCA-Dendrogramm (Anhang A2) fällt unmittelbar auf, dass zwölf einzelne Profile (Fru-2h- $\{a,b,c\}$, Fru-7h- $\{a,b,c\}$, Fru-12h-a, Glu- $\{4h,5h\}$ -c und Glu-9h- $\{a,b,c\}$) vom Rest unabhängig auf höheren Hierarchieebenen gruppiert wurden. Auf einer entsprechend hohen Hierarchieebene wurden im Fall der ESOM-Analyse sieben Profile zugeordnet (Fru-2h- $\{a,b,c\}$, Fru-3h-a, Glu-9h- $\{a,b,c\}$). Grundsätzlich kann ein solches Muster immer zwei Ursachen haben: Zum einen kann das jeweilige Profil ein echter Ausreißer sein und wird daher zu den anderen Profilen distanziert geclus-

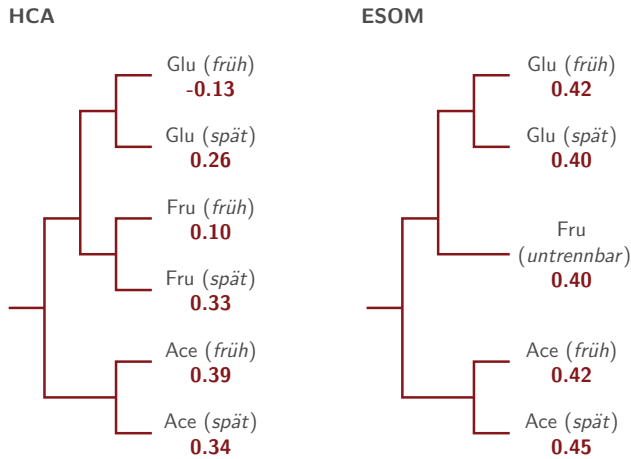


Abbildung 3.8: Vergleichende Darstellung der reduzierten Dendrogramme der metabolischen Profile nach ESOM-Analyse und HCA. Es wurden solche Cluster zusammengefasst, die nach der eingesetzten Kohlenstoffquelle (Fru: Fructose, Glu: Glucose und Ace: Acetat) und der Wachstumsphase diskriminierten (frühe und späte exponentielle Wachstumsphase). Zu jedem Cluster ist die errechnete, durchschnittliche Silhouettenbreite angegeben.

tert. Zum anderen, kann es passieren, dass der Clusteralgorithmus die Ähnlichkeit zweier Profile bzw. zweier Cluster der Struktur der Daten widersprechend bewertet. Um also die Qualität der Clusterprozeduren objektiv beurteilen zu können, war es nötig die Richtigkeit der Zurodnung eines Profils in das jeweilige Cluster quantitativ messen zu können. Hierzu bot sich die Silhouettenbreite als geeignetes Instrument an. Auf Basis von Inter- und Intra-Clusterdistanzen der metabolischen Profile wird ein Wert für jedes ausgewählte Cluster ermittelt, der im Intervall $[-1, 1]$ liegt. Umso höher dieser Wert ist, desto zuverlässiger war die Zuordnung der metabolischen Profile in das jeweilige Cluster. Um also die Silhouettenbreite berech-

nen zu können, wurden in den beiden Dendrogrammen die jeweils größten Cluster ausgewählt, die nach Wachstumsphase und verwendeter Kohlenstoffquelle diskriminierten (Abbildung 3.8). Bezüglich der Wachstumsphase wurde in frühe und späte exponentielle Wachstumsphase unterschieden, wobei das Einsetzen der späten Exponentialphase durch das Clustern nach etwa 4-6 h Kultivierungsdauer bestimmt werden konnte. Im Fall der ESOM-Analyse konnten für die Fructose-Kultivierung keine zwei Cluster gefunden werden, die sich zeitlich in der Exponentialphase abgrenzen ließen.

Grundsätzlich zeigte jedoch die ESOM-Analyse über alle gefundenen Cluster teilweise deutlich höhere Werte für die Silhouettenbreite (0,40 - 0,45) als die HCA (-0,13 - 0,39). Damit bleibt abschließend zu sagen, dass durch das angewandte Prinzip der KNNs die Methode der ESOMs deutlich bessere Ergebnisse bezüglich Clustering und Diskriminierung erlauben, als es durch die HCA möglich ist. Wahrscheinlich ist der kritische Schritt in der HCA die Wahl der Distanzfunktion, die vorab ein detailliertes Wissen über die innere Struktur der Daten erfordert und selbst dann inherent fehlerbehaftet sein kann. Eine solche Wahl findet im Fall der ESOMs nicht statt. Anstattdessen werden die Daten durch das Training des neuronalen Netzes immer wieder über einen nicht-linearen Ansatz aneinander angepasst, was offensichtlich zu robusteren Ergebnissen führte.

3.2 Lichtabhängige Physiologie und Morphologie von *D. shibae*

3.2.1 Populationsdynamik unter fluktuierenden Lichtbedingungen

3.2.1.1 Motivation eines spieltheoretischen Modellierungsansatzes

Eingangs (Kapitel 1.3.2) wurden die photosynthetischen Eigenschaften von *D. shibae* aufgezeigt. Es verfügt über die genetische Ausstattung photosynthetisch aktive Pigmente wie BChl-*a* und Spheroideon zu synthetisieren und es lichtgetrieben unter aeroben Bedingungen zur ATP-Produktion zu nutzen (Wagner-Döbler *et al.*, 2010; Allgaier *et al.*, 2003). Absorptionsspektren aus Zellextrakten deuten weiterhin darauf hin, dass *D. shibae* auf einen Lichtsammelkomplex vom Typ 2 (LH2) verfügt (Biebl *et al.*, 2005; Shiba, 1991). Gleichzeitig konnte gezeigt werden, dass unter Bedingungen, in denen der Organismus Licht ausgesetzt ist, abrupt die Produktion von BChl-*a* einstellt und erst unter dunklen Bedingungen wieder aufzunehmen beginnt (Tomasch *et al.*, 2011; Biebl & Wagner-Döbler, 2006). Dieses Phänomen deutet auf eine augenscheinlich sehr paradoxe Physiologie hin. Es drängt sich die Frage auf, wieso ein Individuum – und ein dazu evolutionär äußerst erfolgreiches – Energie in die Produktion von etwas investiert, was ihm zum Zeitpunkt der Produktion keinen Mehrwert bringt, nur um dann unter Mehrwert schöpfenden Umweltbedingungen, die Produktion dessen wieder einzustellen.

Um sich dieser Frage zu nähern, war zunächst eine Möglichkeit Kulturen von *D. shibae*, die sowohl unter Licht- wie Dunkelheitsbedingungen gewachsen waren, licht- und fluoreszenzmikroskopisch zu untersuchen. Die Fluoreszenz-Excitation wurde mit einem Wellenlän-

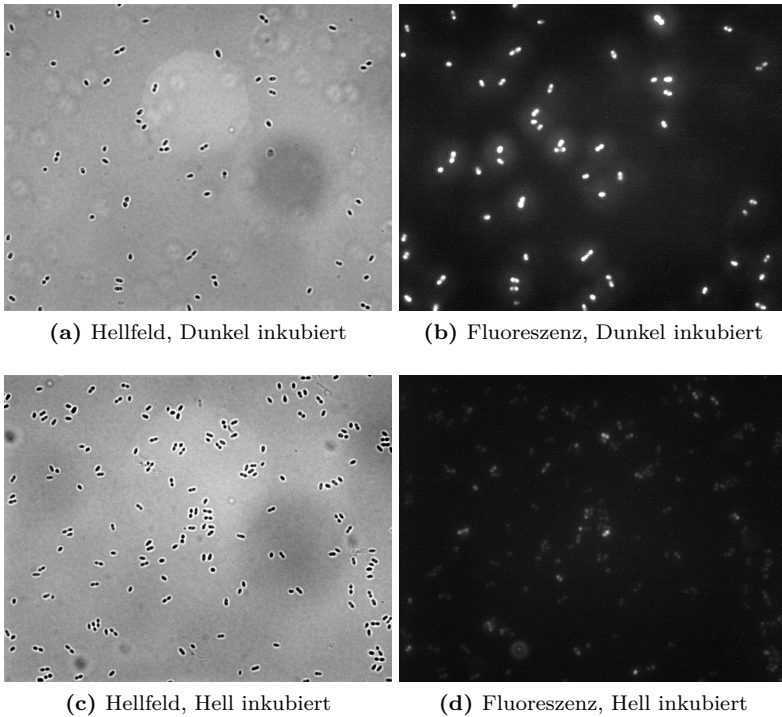


Abbildung 3.9: 100× Vergrößerung von Zellen planktonisch, exponentiell wachsender *D. shibae* Kulturen. a) zeigt Zellen, die unter strikt dunklen Bedingungen wuchsen; c) solche die über 7 h unter Lichteinfluss gewachsen waren. b) und d) sind Fluoreszenzaufnahmen der jeweils linken Abbildung mit $\lambda_{\text{ex}} = (800 \pm 10) \text{ nm}$ und $\lambda_{\text{em}} > 850 \text{ nm}$. Deutlich zu sehen ist, dass die Fluoreszenz in d) niedriger ist als in b). Es ist zudem erkennbar, dass die Helligkeit der Zellen in d) unterschiedlich ist und, dass vereinzelte Zellen bis zu drei voneinander unterscheidbare *helle Flecken* haben.

genbereich $\lambda_{\text{ex}} = (800 \pm 10)$ nm dabei so gewählt, dass damit die Q_y -Bande des LH2 angeregt werden konnte. Die Fluoreszenzemission wurde bei $\lambda_{\text{em}} > 850$ nm gemessen. Die Zellen der dunkel gewachsenen Kultur zeigten unter diesen Bedingungen eine deutliche, und über die Zellen, sehr homogene Fluoreszenzemission. Dies ließ auf ein deutliches Vorhandensein von BChl-*a* schließen (Abbildung 3.9 b). Die im Hellen gewachsene Kultur zeigte hingegen erwartungsgemäß eine deutlich abgeschwächte Fluoreszenz (Abbildung 3.9 d). Bemerkenswerterweise konnte man aber im Gegensatz zu den dunkel kultivierten Individuen hier deutliche Schwankungen in der Helligkeit ausmachen. Es scheint zudem so, dass einzelne Zellen sogar intern heterogen pigmentiert sind und an bis zu drei verschiedenen Stellen deutlich heller leuchten, als der restliche Querschnitt. Diese Beobachtung motivierte die Vermutung, dass *D. shibae* intrazelluläre Strukturen aufweisen könnte, wie sie aus grünen Schwefelbakterien wie z.B. *Chloroflexus aurantiacus* (Allen & Martin, 2007) oder *Chlorobium tepidum* (Pšencík *et al.*, 2004) bekannt sind. Diese Strukturen werden Chlorosomen genannt (Cohen-Bazire *et al.*, 1964) und sind durch eine Lipid-Membran vom umgebenden Cytosol getrennt. Sie enthalten maßgeblich BChl-*a* -Aggregate, aber auch Carotinoide und Quinone und sind der Ort an dem die Photosynthese stattfindet. Weiterhin sind Chlorosomen ellipsoid und können eine Größe von bis zu $(200 \times 100 \times 30)$ nm annehmen.

Hätte also *D. shibae* tatsächlich solche intrazellulären Strukturen in niedriger Anzahl und würde darüber hinaus unter hell-inkubierenden Bedingungen jedes Individuum nicht immer die selbe Anzahl derer tragen, würde dies ein neues Licht auf das eingangs formulierte Paradoxon. Dies mit geeigneten mathematischen Methoden theoretisch zu untersuchen ist Hintergrund des vorliegenden Kapitels. Es wird die Hypothese aufgestellt, dass die Anzahl dieser fluoreszieren-

den, putativ vorkommenden Partikel ein Kulturheterogenitätskriterium definiert und sich in *D. shibae*-Kulturen unter verschiedenen Umweltbedingungen stets eine stabile Mischung der untereinander heterogenen Subpopulationen ausbildet (Kapitel 1.4.1). Ferner wird der Frage nachgegangen, wie sich ein solcher populationsdynamischer Mechanismus auf die Fitness der Gesamtpopulation auswirkt und sich dadurch auch die zum Licht azyklische Produktion von BChl-*a* besser verstehen lässt.

Das im Folgenden entwickelte Modell stellt eine konzeptionelle Erweiterung des Stufen-Matrix Models³ (Caswell, 2001) dar, welches in das Konzept einer kontinuierlichen Replikatordynamik (Eigen, 1971; Schuster & Sigmund, 1983; Cressman, 2003) eingebettet wird. Zunächst werden grundlegende Eigenschaften der Population und deren Zustandsgrößen definiert. Im nächsten Schritt werden die Transitionen zwischen den Stufen der Subpopulationen durch Teilung hergeleitet. In einem weiteren Schritt werden Transitionen durch weitere Effekte beschrieben. Letztendlich werden die Ergebnisse der letzten beiden Schritte zusammengefasst und in die Replikatordynamik eingebettet. Der letzte Schritt der Modellbildung umfasst die Definition der Umwelt und deren Integration in das Modell.

3.2.1.2 Definition von Zustands- und Ereignisraum

Gegeben sei eine Population \mathcal{C} , die sich asexuell durch Teilung ihrer Individuen C_a vermehrt. Grundlegendes Charakteristikum dieser Population ist, dass jedes Individuum eine bestimmte Anzahl sog. *Distribusome* trägt. Der Begriff wird an dieser Stelle neu eingeführt und wurde bewusst abstrakt formuliert, da das im Folgenden entwi-

³ *Engl.*: Stage-Classified Matrix Models. Der Begriff „Stufen-Matrix“ darf in dem Fall nicht mit der oft umgangssprachlich als Stufenmatrix bezeichneten Dreiecksmatrix verwechselt werden.

ckelte Modell ein grundsätzliches Phänomen beschreiben soll. Distribusome zeichnen sich maßgeblich durch drei Eigenschaften aus:

1. Sie kommen in einer sehr niedrigen Anzahl in lebenden Zellen vor und es gibt eine maximale Anzahl, die eine Zelle tragen kann.
2. Zum Zeitpunkt der Zellteilung werden sie an die entstehenden Tochterzellen verteilt (*distribuiert*).
3. Ihre intrazelluläre Anzahl hat einen Einfluss auf deren Fitness, was sich durch eine veränderte Teilungsrate ausdrücken lässt.

Solange insbesondere die erste Eigenschaft gewährleistet ist, stellen Plasmide, Transkriptionsfaktoren, Einschlusskörperchen (Inclusion Bodies), Mitochondrien oder eben Chlorosomen Objekte dar, die sich als Distribusomen beschreiben lassen.

Da grundsätzlich zwei Individuen der gleichen Population eine unterschiedliche Anzahl von Distribusomen tragen können, definiert diese Anzahl ein Heterogenitätskriterium. Daher sei \mathcal{C}^{κ_i} die Subpopulation, deren Individuen $\kappa_i \in \{0, 1, \dots, s\}$ Distribusomen tragen. s definiert die maximale Anzahl der Distribusomen, die ein Individuum tragen kann. Der Einfachheit halber⁴ gilt $\kappa_i = i - 1 \forall i \in [1, s + 1]$. Da alle Subpopulationen \mathcal{C}^{κ_i} untereinander disjunkt sind, gilt für die Gesamtpopulation: $\mathcal{C} = \mathcal{C}^0 \cup \mathcal{C}^1 \cup \dots \cup \mathcal{C}^s$ und $\mathcal{C}^p \cap \mathcal{C}^q = \emptyset \forall p, q \in \{0, 1, \dots, s\}, p \neq q$.

Sei die Anzahl an Individuen der Population \mathcal{C} durch $n = |\mathcal{C}|$ gegeben, sowie die Anzahl der Individuen einer Teilpopulation durch

⁴Es ist für das gesamte Modell unerheblich ob κ_i alle Werte in $[0, s]$ annehmen darf oder eine bestimmte Anzahl von Distribusomen kleiner s nicht definiert ist. Es vereinfacht aber im Folgenden die Notation erheblich und soll daher entsprechend gelten.

$n_i = |\mathcal{C}^{\kappa_i}|$. Alle n_i lassen sich zu dem Vektor $\mathbf{n} \in \mathbb{R}^{s+1}$ zusammenfassen. Entsprechend gilt das Skalarprodukt

$$n = \langle \mathbf{n}, \mathbf{1} \rangle = \sum_{i=1}^{s+1} n_i \quad (3.1)$$

mit $\mathbf{1} = (1, 1, \dots, 1)^T$. Sei weiterhin $x_i = \frac{n_i}{n}$ als der relative Anteil einer Subpopulation \mathcal{C}^{κ_i} an der Gesamtpopulation definiert. Alle x_i lassen sich zu dem Vektor $\mathbf{x} \in \mathbb{R}^{s+1}$ zusammenfassen, für den trivialerweise gilt:

$$\langle \mathbf{x}, \mathbf{1} \rangle = \sum_{i=1}^{s+1} x_i = 1. \quad (3.2)$$

Alle x_i bzw. n_i und ihre zugehörigen Vektoren spannen im Folgenden den Zustandsraum des Modells auf.

Der Ereignisraum des Modells ist als kontinuierliche Zeit $t \in \mathbb{R}_0^+$ definiert. Innerhalb des Ereignisraums unterliegen die Population als ganzes und ihre Subpopulationen und deren Verteilungen dynamischen Veränderungen. Die Zustandsgrößen lassen sich daher auch zeitabhängig als $x_i(t)$ bzw. $n_i(t)$ und $\mathbf{x}(t)$ bzw. $\mathbf{n}(t)$ darstellen.

Grundlegend ist, dass die Menge der κ_i die *Populationsstufen* eines Stufen-Populations (oder im Folgenden Stufen-Matrix) Modells definiert. Ein Individuum einer Subpopulation kann also in eine andere Subpopulation (Stufe) übergehen, indem sich die Anzahl der Distribusome, die es trägt, verändert. Diese Übergänge können sowohl durch Produktion und Degradation der Distribusome als auch durch Zellteilung erfolgen.

3.2.1.3 Rücktransitionen im Stufen-Matrix Modell durch diskretes Wachstum

Am Besten veranschaulichen lässt sich das Stufen-Matrix Modell durch einen Graphen, in dem jede Subpopulation durch einen Knoten repräsentiert wird (Abbildung 3.10). Einzelne Individuen können von einer Stufe in eine andere dann wechseln, wenn für den Wechsel im Graphen eine gerichtete Kante existiert. Der Anteil der Individuen, die in einem Zeitabschnitt $t \rightarrow t + 1$ von Stufe j nach Stufe i wechseln ist durch den Koeffizienten $a_{i,j}$ gegeben. Ordnet man die einzelnen Koeffizienten in einer Matrix $\mathbf{A} \in \mathbb{R}^{s+1 \times s+1}$ an und berücksichtigt man, dass die Verteilung der Subpopulationen zu einem Zeitpunkt t durch $\mathbf{x}(t)$ gegeben ist, erhält man für die zeitdiskrete Dynamik des Modells den Zusammenhang:

$$\mathbf{x}(t + 1) = \mathbf{A}\mathbf{x}(t). \quad (3.3)$$

Nun soll der Zeitabschnitt $t + 1$ genau einer Verdopplungszeit über alle Subpopulationen entsprechen. Dann sind die Koeffizienten $a_{i,j}$ davon abhängig, wie sich die Distribusome während der Zellteilung auf die beiden Tochterzellen verteilen. Anhand des Schicksals eines Individuums aus \mathcal{C}^3 über zwei Verdopplungszeiten ist dies in Abbildung 3.11 veranschaulicht. Es wird zunächst der allgemeine Fall angenommen, dass die Nachkommen eines Individuums grundsätzlich unterscheidbar sind, wie es z.B. bei einem Populationswachstum durch Knospung der Fall ist. Durch Teilung des Individuums verteilt es seine Distribusome an die entstehenden Tochterzellen. Dies ist die fundamentale Annahme des gesamten hier entwickelten Modells. Die Matrix ${}^+\mathbf{Q} \in \mathbb{R}^{s+1 \times s+1}$ definiert die jeweiligen Übergangswahrscheinlichkeiten ${}^+q_{i,j}$, mit der ein Individuum von einer Populations-

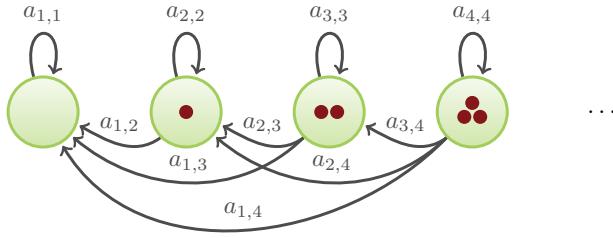


Abbildung 3.10: Gerichteter Graph eines Stufen-Matrix Modells. Jeder Knoten repräsentiert eine Stufe der Population; jede Kante einen möglichen Übergang eines Individuums von einer Stufe in eine andere. Die Fraktion der Individuen, die während $t \rightarrow t+1$ von einer Stufe j in eine andere i wechselt, ist mit den jeweiligen Koeffizienten $a_{i,j}$ angegeben. Rot illustriert sind die Distribusome, deren Anzahl die jeweilige Populationsstufe charakterisiert.

stufe j in eine andere i übergeht⁵. Analog definiert ${}^{-}\mathbf{Q} \in \mathbb{R}^{s+1 \times s+1}$ die Übergangswahrscheinlichkeiten für das zweite entstehende Individuum.

${}^{+}\mathbf{Q}$ hat die folgenden Eigenschaften:

$$0 \leq {}^{+}q_{i,j} \leq 1 \quad \forall \quad i, j \in \{1, 2, \dots, s+1\} \quad (3.4)$$

$$\sum_{i=1}^{s+1} {}^{+}q_{i,j} = 1 \quad \forall \quad j \in \{1, 2, \dots, s+1\} \quad (3.5)$$

$${}^{+}q_{i,j} = 0 \quad \forall \quad i > j. \quad (3.6)$$

Die Ungleichung 3.4 stellt sicher, dass die einzelnen Koeffizienten im-

⁵Üblicherweise werden Übergangs- oder Transitionsmatrizes in der transponierten Form angegeben, um Übergänge $i \rightarrow j$ zu beschreiben. Da die transponierte Form der Matrix im weiteren Verlauf bestimmte Herleitungen unnötig verkomplizieren würde, wird hier darauf verzichtet und ${}^{+}\mathbf{Q}^T$ dann verwendet, wenn konkret allgemeine Eigenschaften von Transitionsmatrizes gezeigt werden sollen.

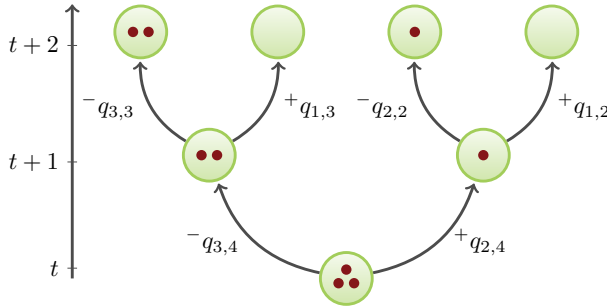


Abbildung 3.11: Illustration des Schicksals eines Individuums der Subpopulation \mathcal{C}^3 über zwei Verdopplungszeiten $t \rightarrow t+1 \rightarrow t+2$. Durch Teilung verteilen sich die Distribusome des Individuums auf die Tochterzellen, was den Übergang einer Populationsstufe in eine andere determiniert. Die Wahrscheinlichkeit mit der ein Individuum von einer Populationsstufe j in eine andere i übergeht, ist durch die Übergangswahrscheinlichkeiten $+q_{i,j}$ bzw. $-q_{i,j}$ definiert.

mer eine Wahrscheinlichkeit ausdrücken. Ferner ist die Wahrscheinlichkeit von einem bestimmten Zustand j in irgendeinen Zustand i überzugehen durch die Spaltensumme gegeben und muss immer 1 ergeben (Gleichung 3.5). Da weiterhin ein Übergang eines Individuums von einer Subpopulation $\mathcal{C}^a \rightarrow \mathcal{C}^b$ mit $b > a$ eine Vermehrung der Distribusome bedeuten würde, sind solche Übergänge hier unzulässig. Dies wird durch Gleichung 3.6 gewährleistet. Damit ist auch festgelegt, dass $+Q$ immer eine obere Dreiecksmatrix bildet. Alle drei Eigenschaften gelten in vollkommener Analogie auch für $-Q$.

Letztendlich sind die Übergangskoeffizienten in $-Q$ und $+Q$ voneinander nicht unabhängig. Nach der Teilung muss die Anzahl, der auf die Tochterzellen verteilten Distribusome derjenigen der Mutterzelle entsprechen. Damit ergibt sich, dass der Übergang einer Tochterzelle von $\mathcal{C}^a \xrightarrow{+} \mathcal{C}^b$ festlegt, dass die zweite Tochterzelle immer im Übergang $\mathcal{C}^a \xrightarrow{-} \mathcal{C}^{a-b}$ gebildet wird. Dementsprechend muss für diesen

Übergang die selbe Wahrscheinlichkeit gelten:

$${}^+q_{i,j} = {}^-q_{j-i+1,j} \quad \forall \quad i \leq j \quad (3.7)$$

Es wird nun angenommen, dass die Übergangswahrscheinlichkeiten einzelner Individuen bei einer hinreichenden Populationsgröße den Anteilen der Subpopulationen entsprechen, die während eines Verdopplungszyklus in die jeweilig anderen Subpopulationen übergehen. Damit ergibt sich die folgende Definition.

Definition 1 *Die Häufigkeitsverteilung von Individuen in Subpopulationen zu einem Zeitpunkt t ist gegeben durch $\mathbf{x}(t)$. Nachdem sich jedes Individuum genau einmal teilt, ergibt sich die neue Häufigkeitsverteilung aus*

$$\mathbf{x}(t+1) = \frac{1}{2}({}^- \mathbf{Q} + {}^+ \mathbf{Q}) \cdot \mathbf{x}(t). \quad (3.8)$$

Die Verteilung nach r Generationszeiten entspricht somit

$$\mathbf{x}(t+r) = \left[\frac{1}{2}({}^- \mathbf{Q} + {}^+ \mathbf{Q}) \right]^r \cdot \mathbf{x}(t). \quad (3.9)$$

Letztendlich lässt sich das hier entworfene Matrix-Stufen Modell ohne Umstände in das Konzept der Markov Ketten (Grinstead & Snell, 1997) einbetten, weshalb sich eine Reihe relevanter Sätze aus diesem Bereich unmittelbar auf dieses Modell anwenden lassen.

Korollar 1 *Die transponierten Matrizes ${}^- \mathbf{Q}^T$, ${}^+ \mathbf{Q}^T$ und $\frac{1}{2}({}^- \mathbf{Q} + {}^+ \mathbf{Q})^T$ sind Transitionsmatrizes einer Markov Kette.*

Beweis:

1. ${}^- \mathbf{Q}$, ${}^+ \mathbf{Q}$ sind aufgrund ihrer Eigenschaften 3.4 und 3.5 eine Markov Transitionsmatrix.

2. Eigenschaft 3.4 gilt auch für $\frac{1}{2}(-\mathbf{Q} + +\mathbf{Q})$:

$$0 \leq \frac{1}{2}(+q_{i,j} + -q_{i,j}) \leq 1 \quad \forall i, j \in \{1, 2, \dots, s+1\}.$$

3. Eigenschaft 3.5 gilt auch für $\frac{1}{2}(-\mathbf{Q} + +\mathbf{Q})$:

$$\sum_{i=1}^{s+1} \frac{1}{2}(+q_{i,j} + -q_{i,j}) = \frac{1}{2} \left(\sum_{i=1}^{s+1} +q_{i,j} + \sum_{i=1}^{s+1} -q_{i,j} \right) = 1$$

$$\forall j \in \{1, 2, \dots, s+1\}.$$

■

Korollar 2 Die transponierten Matrizes $-\mathbf{Q}^T$, $+\mathbf{Q}^T$ und $\frac{1}{2}(-\mathbf{Q} + +\mathbf{Q})^T$ sind Transitionsmatrizes absorbierender Markov Ketten.

Beweis: Eine Transitionsmatrix, die in Dreiecksform vorliegt (Gleichung 3.6 bzw. Gleichung 3.19), wirkt immer absorbierend. ■

Das Korollar 2 beschreibt das Langzeitverhalten des hier entwickelten Modells und deutet bereits eine weitreichende Konsequenz an. Es besagt letztendlich dass in hinreichend langer Zeit der Term $[\frac{1}{2}(-\mathbf{Q} + +\mathbf{Q})]^r$ gegen eine stabile (absorbierende) Matrix strebt und somit auch die Häufigkeitsverteilung $\mathbf{x}(t+r)$ gegen eine stabile, sich ohne äußere Einflüsse nicht mehr ändernde, Häufigkeitsverteilung \mathbf{x}^* strebt:

$$\lim_{r \rightarrow \infty} \mathbf{x}(t+r) = \mathbf{x}^*. \quad (3.10)$$

Was hier trivialerweise gezeigt werden konnte, leitet sich auch in der Form aus Abbildung 3.10 ab: Da es keine Transitionen $\mathcal{C}^a \rightarrow \mathcal{C}^b$ mit $b > a$ gibt bzw. deren Übergangswahrscheinlichkeit $p_{i,j} = 0$ ist, strebt die gesamte Population immer der Subpopulation \mathcal{C}^0 entgegen. Im Folgenden werden noch Fälle entwickelt, die weniger trivial sind.

3.2.1.4 Übergang des Stufen-Matrix Modells in kontinuierliches Wachstum

Das zeitdiskrete Modell in Gleichung 3.9 soll nun auf einen kontinuierlichen Ereignisraum übertragen werden. Hierzu ist es zunächst nötig das Modell nicht für die Häufigkeitsverteilungen der Subpopulationen, sondern die Anzahl absoluter Individuen zu übertragen. Ferner wird nun angenommen, dass jede Subpopulation \mathcal{C}^{i-1} ($i \in [1, s + 1]$) mit einer individuellen, zeitinvarianten Wachstumsrate μ_i wächst. Die Menge aller Wachstumsraten lässt sich als Vektor $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_{s+1})^T$ ausdrücken. Ohne Berücksichtigung von Transitionen zwischen den einzelnen Subpopulationen ergibt sich das ungekoppelte Wachstumsmodell in Matrix-Schreibweise nach

$$\frac{d\mathbf{n}(t)}{dt} = \boldsymbol{\mu}^D \mathbf{n}(t). \quad (3.11)$$

Hierbei repräsentiert $\boldsymbol{\mu}^D \in \mathbb{R}^{s+1 \times s+1}$ die diagonalisierte Form der Wachstumsraten

$$\boldsymbol{\mu}^D = \mathbf{I} \cdot \boldsymbol{\mu} = \begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mu_{s+1} \end{pmatrix} \quad (3.12)$$

mit $\mathbf{I} \in \mathbb{R}^{s+1 \times s+1}$ als der Identitätsmatrix. Zur Integration der Transitionen in Gleichung 3.11 werden nun drei Schritte benötigt, die an die Entwicklung eines Markov Geburt-Tod-Modells angelehnt sind (Novak, 2006):

1. Die Anzahl der Individuen $\mathbf{n}^*(t) \cdot dt$, die sich in allen Subpopulationen im Zeitraum dt verdoppeln würde ist

$$\mathbf{n}^*(t)dt = \boldsymbol{\mu}^D \cdot \mathbf{n}(t) \cdot dt.$$

Sie werden aus dem Bilanzraum des Modells entfernt.

2. Genau diese Anzahl von Individuen verdoppelt und verteilt sich entsprechend der Transitionsmatrizes ${}^+\mathbf{Q}$ und ${}^-\mathbf{Q}$ auf die Subpopulationen. Dabei treten die Individuen wieder in den Bilanzraum ein. Die Änderungen der Größen der Subpopulationen sind gegeben durch

$$\mathbf{n}^*(t + dt) \cdot dt = ({}^+\mathbf{Q} + {}^-\mathbf{Q}) \cdot \mathbf{n}^*(t) \cdot dt.$$

3. Berücksichtigt man beide Prozesse ergibt sich für die Nettoänderungen $d\mathbf{n}(t)$:

$$d\mathbf{n}(t) = -\mathbf{n}^*(t) \cdot dt + \mathbf{n}^*(t + dt) \cdot dt.$$

Definition 2 *Die kontinuierliche Dynamik einer Population von Individuen, die sich durch ihre Teilung auf verschiedene Subpopulationen umverteilen, ist gegeben durch*

$$\frac{d\mathbf{n}(t)}{dt} = ({}^-\mathbf{Q} + {}^+\mathbf{Q} - \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t). \quad (3.13)$$

Dass in dem Modell bisher keine Produktion oder Degradation der Distribusome berücksichtigt wurde, führt unmittelbar zu dem folgenden Satz.

Satz 1 *Die Gesamtanzahl der Distribusome in einer Population \mathcal{C} , die sich gemäß Definition 2 verhält ist in $[t, \infty)$ stets konstant:*

$$\frac{d\langle \mathbf{n}(\mathbf{t}), \begin{pmatrix} 0 \\ 1 \\ \dots \\ s \end{pmatrix} \rangle}{dt} = 0. \quad (3.14)$$

Der Beweis dieses Satzes findet sich in Anhang 4.2.

3.2.1.5 Stochastische Zufallsverteilung

Im Weiteren wird für die Verteilung der Distribusome die Annahme getroffen, dass bei dem Ereignis der Teilung eines Individuums, jedes einzelne Distribusom mit der gleichen Wahrscheinlichkeit r beim Übergang $\mathcal{C}^a \xrightarrow{+} \mathcal{C}^b$ an das Tochterindividuum übertragen wird. Für die Gesamtwahrscheinlichkeit, dass dieser Übergang stattfinden kann dann die Binomialverteilung $B(b|r, a)$ angewendet werden. Analog ist die Wahrscheinlichkeit eines beliebigen Übergangs $\mathcal{C}^a \xrightarrow{-} \mathcal{C}^b$ gegeben durch $B(b|1-r, a)$:

$${}^+p_{i,j} = B(i-1 | r, j-1) = \binom{j-1}{i-1} r^{i-1} (1-r)^{j-i} \quad (3.15)$$

$${}^-p_{i,j} = B(i-1 | 1-r, j-1) = \binom{j-1}{i-1} (1-r)^{i-1} r^{j-i}. \quad (3.16)$$

Es soll von nun an angenommen werden, dass bei der Verteilung der Distribusome beide Tochterindividuen ununterscheidbar sind. Daher wird $r = 0,5$ gesetzt und die letzte Gleichung vereinfacht sich zu:

$${}^+p_{i,j} = B(i-1 | 1/2, j-1) = (1/2)^{j-1} \cdot \binom{j-1}{i-1} \quad (3.17)$$

$${}^-p_{i,j} = B(i-1 | 1/2, j-1) = (1/2)^{j-1} \cdot \binom{j-1}{i-1}. \quad (3.18)$$

Konsequenterweise ist aus den den letzten beiden Gleichungen unmittelbar zu sehen, dass ${}^+p_{i,j} = {}^-p_{i,j} \forall i, j \in \{1, 2, \dots, s+1\}$ und somit ${}^-Q = {}^+Q = Q$ ist. Für $r = 0,5$ und $s = 3$ ist Q beispielsweise:

$$Q = \begin{pmatrix} 1 & 1/2 & 1/4 & 1/8 \\ 0 & 1/2 & 1/2 & 3/8 \\ 0 & 0 & 1/4 & 3/8 \\ 0 & 0 & 0 & 1/8 \end{pmatrix}. \quad (3.19)$$

3.2.1.6 Vorwärtstransitionen im Stufen-Matrix Modell durch Produktion

Bisher wurde nur die Verteilung von Distribusomen durch Wachstum bzw. Teilung der Individuen hergeleitet. Im Folgenden soll auch deren Produktion berücksichtigt werden. Eine Degradation wird weiterhin außer Acht gelassen. Wieder wurde hierzu ein Stufen-Matrix Modell erzeugt (Abbildung 3.12). Die Koeffizienten $b_{i,j}$ geben an, mit welcher Wahrscheinlichkeit in einem bestimmten Zeitintervall t_b ein Distribusom produziert wird und somit welche Fraktion der Individuen in t_b den Übergang $\mathcal{C}^{j-1} \rightarrow \mathcal{C}^{i-1}$ (also von der Subpopulation deren Individuum $j - 1$ Distribusomen tragen zu der, deren Individuen $i - 1$ Distribusomen tragen) vollziehen. Die einzelnen Koeffizienten $b_{i,j}$ formen die Matrix $\mathbf{B} \in \mathbb{R}^{s+1 \times s+1}$, die damit die folgende Form annimmt:

$$\mathbf{B} = \begin{pmatrix} 1 - b_{2,1} & 0 & \dots & 0 & 0 \\ b_{2,1} & 1 - b_{3,2} & \dots & 0 & 0 \\ 0 & b_{3,2} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 - b_{s+1,s} & 0 \\ 0 & 0 & \dots & b_{s+1,s} & 1 \end{pmatrix} \quad (3.20)$$

Demnach gelten wieder die Eigenschaften 3.4 und 3.5. Ferner ist \mathbf{B}^T wieder eine Transitionsmatrix einer absorbierenden Markov Kette.

\mathbf{B} wird nun unmittelbar in die kontinuierliche Dynamik (Gleichung 3.11) integriert. Dies wird dadurch vereinfacht, dass die Übergangswahrscheinlichkeitskoeffizienten $b_{i+1,i}$ in Produktionsraten überführt werden. Hierzu wird die Produktionsmatrix $\mathbf{P} \in \mathbb{R}^{s+1 \times s+1}$ eingeführt. Sie nimmt die selbe Form wie \mathbf{B} an. Die Koeffizienten $p_{i+1,i}$ geben die Produktionsrate für die Produktion von $i \rightarrow i + 1$ Distri-

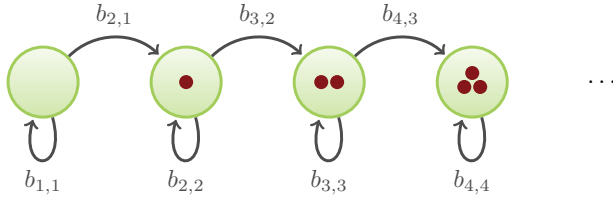


Abbildung 3.12: Gerichteter Graph eines Stufen-Matrix Modells. In diesem Modell werden keine Rückwärtstransitionen berücksichtigt. Die Wahrscheinlichkeiten $b_{i,j}$ geben an mit welcher Wahrscheinlichkeit ein Individuum, während $t \rightarrow t + 1$ von einer Stufe j in eine andere i wechselt. Die Stufen der einzelnen Subpopulationen sind wieder durch die Anzahl der Distribusome (Rot) illustriert.

busomen als Vielfaches der jeweiligen Wachstumsrate μ_i an. Damit ergibt sich in vollkommener Analogie zur Herleitung der zeitkontinuierlichen Rückwärtstransitionen in Kapitel 3.2.1.4, die folgende Definition:

Definition 3 Die kontinuierliche Dynamik einer Population von Individuen, die sich durch Produktion der Distribusome auf verschiedene Subpopulationen umverteilen ist gegeben durch

$$\frac{d\mathbf{n}(t)}{dt} = (\mathbf{P} - \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t). \quad (3.21)$$

$p_{i+1,i} \cdot \mu_i$ ist die Rate mit der die Individuen aus einer Subpopulation \mathcal{C}^{i-1} in \mathcal{C}^i übergehen.

Daraus, dass es bei diesem Prozess lediglich zu einer Umverteilung der Individuen kommt und kein Nettowachstum berücksichtigt wurde, ergibt sich der folgende Satz.

Satz 2 Die Gesamtanzahl der Individuen in einer Population \mathcal{C} , die

sich gemäß Definition 3 verhält, ist in $[t, \infty)$ stets konstant:

$$\frac{d \sum_i n_i(t)}{dt} = 0 \quad (3.22)$$

Der Beweis dieses Satzes findet sich in Anhang 4.2.

Da angenommen wird, dass die Produktion und Verteilung der Distribusome zwei Vorgänge sind, die sich störungsfrei überlagern, folgt daraus unmittelbar das Differentialgleichungssystem, das beide Vorgänge vereint:

$$\frac{d\mathbf{n}(t)}{dt} = (-\mathbf{Q} + {}^+\mathbf{Q} + \mathbf{P} - 2 \cdot \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t). \quad (3.23)$$

Berücksichtigt man zudem, dass die Verteilung der Distribusome einer symmetrischen Binomialverteilung folgt, liefert das die folgende Definition:

$$\frac{d\mathbf{n}(t)}{dt} = (2 \cdot (\mathbf{Q} - \mathbf{I}) + \mathbf{P}) \boldsymbol{\mu}^D \mathbf{n}(t). \quad (3.24)$$

3.2.1.7 Einbettung in die Replikatorgleichung

Der Matrixterm in Gleichung 3.24 beschreibt letztendlich die *Fitness* $\mathbf{F} \in \mathbb{R}^{s+1 \times s+1}$ der Subpopulationen:

$$\mathbf{F} = (2 \cdot (\mathbf{Q} - \mathbf{I}) + \mathbf{P}) \boldsymbol{\mu}^D. \quad (3.25)$$

Damit lässt sich Gleichung 3.24 unmittelbar in die Replikatordynamik überführen (Kapitel 1.4.3), die in einer allgemeinen und vereinfachten Form

$$\frac{dx_i(t)}{dt} = x_i \cdot ((\mathbf{F}\mathbf{x})_i - \mathbf{x}^T \mathbf{F} \mathbf{x}) \quad (3.26)$$

lautet. Es fällt auf, dass die Replikatorgleichung auf der Häufigkeitsverteilung \mathbf{x} angewandt wird. Tatsächlich liefert dieses Modell

keine Aussagen darüber, wie sich Populationen in absoluten Zahlen entwickeln. Es untersucht lediglich wie sich im zeitlichen Verlauf die Verhältnisse der Populationen zueinander verändern. Dies erklärt auch die Form der Gleichung: $(\mathbf{F}\mathbf{x})_i = \sum_j f_{i,j} \cdot x_j$ ist die individuelle Fitness der Individuen der i -ten Subpopulation, während $\mathbf{x}^T \mathbf{F} \mathbf{x} = \sum_i (\mathbf{F}\mathbf{x})_i \cdot x_i$ die Gesamtfitness der Population beschreibt. Somit ist in diesem Modell nur die Abweichung der individuellen Fitness zur Gesamtfitness relevant (Cressman, 2003; Hofbauer & Sigmund, 2003).

3.2.1.8 Integration von Umweltreizen und Simulation der *Roseobacter* Hypothese

In das hier entwickelte Modell wird nun ein Umweltreiz integriert, der durch die Funktion $\mathcal{L}(t) \in [0, 1]$ abgebildet wird. Ein Wert von 1 bedeutet, dass der Umweltreiz in seiner vollen Ausprägung vorhanden ist, ein Wert von 0, dass seine Ausprägung hingegen minimal ist. \mathcal{L} nimmt insofern auf das Modell Einfluss, als dass es sowohl die Wachstumsraten, wie auch die produktionsbedingten Transitionsraten variiert. Vereinfacht sollen diese beide Parameter linear vom Wert von \mathcal{L} abhängen:

$$\boldsymbol{\mu}(\mathcal{L}) = {}^0\boldsymbol{\mu} + \Delta\boldsymbol{\mu} \cdot \mathcal{L} \quad (3.27)$$

$$\mathbf{P}(\mathcal{L}) = {}^0\mathbf{P} \cdot (1 - \mathcal{L}) \quad (3.28)$$

Das heisst, dass es eine Basiswachstumsrate ${}^0\mu_i$ gibt, die sich bei einsetzendem Umweltreiz linear um $\Delta\mu_i \cdot \mathcal{L}$ verschiebt. Analog gibt es eine produktionsabhängige Basistransitionsrate ${}^0\mathbf{P}$, die mit steigendem Umwelteinfluss linear auf 0 fällt.

Das bis hierher entwickelte Modell lässt sich nun unmittelbar auf die in Kapitel 3.2.1.1 gemachten Annahmen übertragen. Die Distri-

busome nehmen die Rolle der putativ vorhergesagten Chlorosom-ähnlichen Strukturen ein. Ein Individuum kann hierbei maximal $s = 3$ Distribusome tragen. Die Population der *D. shibae* Individuen zerfällt damit in 4 Teilpopulationen $\mathcal{C} = \mathcal{C}^0 \cup \mathcal{C}^1 \cup \mathcal{C}^2 \cup \mathcal{C}^3$. Die Verteilung der Distribusome erfolgt gemäß einer symmetrischen Binomialverteilung (Gleichung 3.17). Diese Population wird in der folgenden Simulation dem Umweltreiz „Licht“ ausgesetzt. Für die übrigen Parameter können für den Moment nur realistische Annahmen getroffen werden: Die lichtunabhängige Basiswachstumsrate wird für alle Subpopulationen auf $0.2 \cdot 1/h$ gesetzt: ${}^0\mu = (0, 2; 0, 2; 0, 2; 0, 2)^T \cdot 1/h$. Da nun die Annahme getroffen wird, dass der lichtabhängige Zuwachs der Wachstumsrate umso größer sein soll je mehr Distribusome das Individuum trägt, wird ${}^\Delta\mu = (0; 0, 02; 0, 035; 0, 045)^T \cdot 1/h$ gesetzt. Für die Produktionsgeschwindigkeit der Distribusome wird angenommen, dass die Produktion eines ersten Distribusoms schneller sein soll, als die eines Zweiten, die wiederum schneller ist als die eines Dritten. Damit⁶ gilt $p_{1,0} > p_{2,1} > p_{3,2}$.

Mit diesen Parametern wurden drei Simulationen des Differentialgleichungssystems 3.26 durchgeführt. In jeder dieser Simulationen nahm der Umweltreiz einen anderen zeitlichen Verlauf ein:

1. $\mathcal{L}(t) = 0$ (Wachstum ohne Licht)

2. $\mathcal{L}(t) = 1$ (Wachstum mit Licht)

3. $\mathcal{L}(t) = \begin{cases} 0 & \forall \quad n \cdot \frac{T}{2} < t < (n+1) \cdot \frac{T}{2}; \quad n \in 0, 2, 4, \dots \\ 1 & \forall \quad n \cdot \frac{T}{2} < t < (n+1) \cdot \frac{T}{2}; \quad n \in 1, 3, 5, \dots \end{cases}$
(Wachstum unter fluktuierenden Lichtbedingungen).

⁶Tatsächlich lässt sich dieser Umstand nur dann so leicht beschreiben, wenn alle ${}^\Delta\mu_i$ im Vergleich zu ${}^0\mu_i$ sehr klein sind. Da aber im weiteren Verlauf sowieso nur angenommen wird, dass $\mathcal{L} = \{0, 1\}$, ist diese Annahme vollkommen störungsfrei.

Der letzte der drei Verläufe beschreibt eine Rechteckschwingung des Umweltreizes mit einer Periodendauer von $T = 24$ h und einer Amplitude von $\{0, 1\}$. Als Anfangsbedingung der Populationsverteilung wurde jeweils $\mathbf{x}(0) = (0, 97; 0, 01; 0, 01; 0, 01)^T$ gewählt, was einer fast homogenen Population mit 0 Distribusomen je Individuum entspricht. Die Simulationen wurden in R unter Zuhilfenahme des Paketes `deSolve`, welches zum numerischen Lösen von Differentialgleichungssystemen geeignet ist, durchgeführt.

Abbildung 3.13, links oben, zeigt den zeitlichen Verlauf der Simulation ohne Einfluss von Licht. Offensichtlich streben die Verhältnisse der Subpopulationen in dieser Simulation einem Fixpunkt entgegen. Eine Simulation über 10.000 h ergibt (auf drei Stellen gerundet) die Verteilung $(\mathbf{x}^*)^T = (0, 276; 0, 361; 0, 299; 0, 065)$. Bei diesem Ergebnis handelt es sich um ein Nash-Gleichgewicht. Dies lässt sich weiterhin leicht zeigen, da $(\mathbf{F}\mathbf{x}^*)_i = (\mathbf{x}^*)^T \mathbf{F}\mathbf{x}^* \quad \forall i \in [1, s + 1]$ erfüllt ist. Da das aufgestellte Modell ein symmetrisches evolutionäres Spiel ist, findet außerdem der Satz von Bishop & Cannings (1978) Anwendung, wonach dieses Gleichgewicht nicht nur ein evolutionär stabiler Zustand (ESS), sondern auch der einzige innere ESS ist (Kapitel 1.4.3). Die Konsequenz daraus ist, dass für alle Anfangsbedingungen \mathbf{x}_0 mit $(x_0)_i \neq 0$ und $\sum_i (x_0)_i = 1 \quad \forall i \in [1, s + 1]$ die Verteilung den Zustand \mathbf{x}^* annehmen wird.

Der zeitliche Verlauf in der Simulation unter konsequent hellen Bedingungen (Abbildung 3.13, rechts oben) zeigt keine überraschenden Effekte und strebt asymptotisch den Fixpunkt $(\mathbf{x}^*)^T = (1; 0; 0; 0)$ an: Alle Subpopulationen, die mindestens ein Distribusom tragen extingieren zeitlich, was anschaulich eine Folge der inhibierten Produktion der Distribusomen ist.

Letztendlich beschreibt die Simulation unter Hell-Dunkel-Zyklen eine Mischform zwischen den beiden, zuletzt genannten Extremfäl-

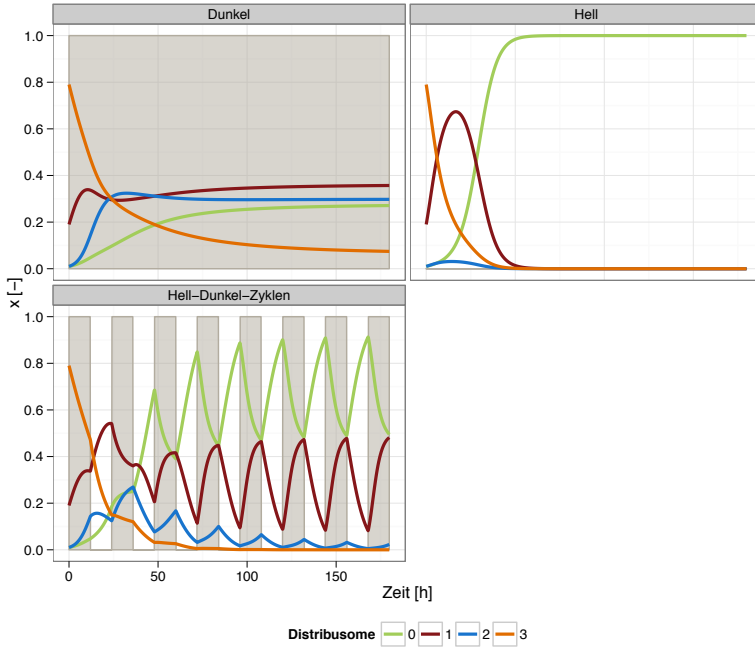


Abbildung 3.13: Dargestellt sind die numerischen Simulationen des Differentialgleichungssystems 3.24 nach Einbettung in die Replikatorgleichung. Die zeitlichen Verläufe über 180 h beschreiben die Subpopulationen mit jeweils keinem (Grün), einem (Rot), zwei (Blau) und drei (Orange) Distribusomen. Grau eingefärbt sind die Intervalle, in denen der Umweltparameter Licht auf $\mathcal{L}(t) = 0$ gesetzt wurde. Links oben ist der simulierte Verlauf einer Population gezeigt, die Dunkelheit ausgesetzt war. Im Langzeitverhalten streben die einzelnen Subpopulationen dem stabilen Gleichgewicht $\mathbf{x}^T = (0,078; 0,215; 0,229; 0,478)$ entgegen. Rechts oben ist der theoretische Verlauf einer Population, die dauerhaft Belichtung ausgesetzt war gezeigt. Die einzelnen Subpopulationen streben stabil gegen $\mathbf{x}^T = (1; 0; 0; 0)$. Unten links ist der Verlauf unter fluktuierenden Lichtbedingungen gezeigt. Die Periodendauer beträgt $T = 24$ h. x_3 und x_4 streben langfristig gegen 0. x_1 und x_2 oszillieren hingegen langfristig stabil. Im letzten Übergang $\mathcal{L} : 1 \rightarrow 0$ zu $t = 168$ h ist $\mathbf{x}^T = (0,913; 0,082; 0,005; 0)$; im letzten Übergang $\mathcal{L} : 0 \rightarrow 1$ zu $t = 180$ h ist $\mathbf{x}^T = (0,494; 0,482; 0,024; 0)$.

len (Abbildung 3.13, unten links). Nach einer Adaptionphase über etwa 100 h beginnt das System asymptotisch stabil zu oszillieren. Eine Simulation über 10.000 h (Abbildung nicht gezeigt, Daten auf drei Stellen gerundet) zeigt für den Übergang Hell zu Dunkel die Verteilung der Subpopulationen $(\mathbf{x}^*)^T = (0,908; 0,092; 0; 0)$ und im Übergang Dunkel zu Hell $(\mathbf{x}^*)^T = (0,516; 0,484; 0; 0)$. Wenn das System aber auch zwischen den Zuständen der Extremfälle oszilliert und letztendlich die Mitglieder der Subpopulationen, die zwei bzw. drei Distribusome tragen asymptotisch extingieren, kommt es bei den gewählten Parametern nie zu einer vollständigen Extinktion der Subpopulation mit einem oder keinem Distribusom.

Anhand dieser Simulation lässt sich letztendlich das folgende Prinzip ableiten: In den dunklen Phasen gewinnt die Subpopulation mit einem Distribusom so große Anteile, dass beim Übergang in die helle Phase die Gesamtpopulation eine erhöhte Nettowachstumsrate hat, als eine reine Kultur deren Mitglieder keine Distribusome tragen. Freilich hängt der Zustand des Systems an den Übergängen von den konkret gewählten Parametern und der Dauer bzw. einer Stochastik der Hell- und Dunkelphasen ab.

3.2.1.9 Experimentelle Modellverifikation

Die im letzten Abschnitt aufgestellte Hypothese sollte nun anhand der eingangs beschriebenen Physiologie von *D. shibae* verifiziert werden. Hierzu sollten die populationsdynamischen Zustandsgrößen ermittelt und die in Gleichung 3.24 aufgeführten Modellparameter experimentell bestimmt werden. Ein Experiment, welches sich grundsätzlich hierzu eignen würde, umfasst optimalerweise das planktonische Wachstum einer *D. shibae* Kultur und ermöglicht zeitaufgelöst das Schicksal einzelner Individuen der Gesamtpopulation zu verfolgen. Die Kultur wird hierbei konstanten Umweltparametern

wie Nährstoffkonzentration, Sauerstoffpartialdruck oder Temperatur unter definierten Hell-Dunkel-Zyklen ausgesetzt.

Eine technische Möglichkeit Analysen auf Einzellzebene durchzuführen stellen fluoreszenzmikroskopische Zeitrafferaufnahmen, immobilisiert wachsender Kulturen dar. Insbesondere die messbare Fluoreszenz des BChl-*a* in *D. shibae* unterstützt eine solch bildgebendes Verfahren. Wenn die Zeitraffermikroskopie aber auch in den vergangenen Jahren eine stete Entwicklung und Verbesserungen erfahren hat und in den Fokus ähnlicher Fragestellungen gerückt war (Bergmiller *et al.*, 2011; Wang & Bodovitz, 2010; Locke & Elowitz, 2009; Young *et al.*, 2012), stehen einem solchen Versuchsaufbau im vorliegenden Fall maßgeblich drei Umständen entgegen:

- Zum einen sind *D. shibae* Individuen bei einer durchschnittlichen Größe von 0,3 - 1,0 μm mit konventionellen Licht- und Fluoreszenzmikroskopen nur noch sehr eingeschränkt mikroskopierbar. Insbesondere sich teilende von bereits geteilten Zellen mit hinreichender Zuverlässigkeit unterscheiden zu können, ist aber zwingend nötig, um die Verdopplungsraten zu bestimmen.
- Zum andere erfordert die Mikroskopie, das Objekt in zumindest regelmäßigen Abständen Licht hoher Intensität auszusetzen. Da aufgrund vorheriger Arbeiten (Biebl & Wagner-Döbler, 2006; Tomasch *et al.*, 2011) jedoch davon auszugehen war, dass insbesondere das lichtabhängige Stoppen der BChl-*a* Produktion bereits bei sehr kurzer Licht-Exposition ausgelöst wird, würde dies mit einer massiven Störung der eigentlichen Beobachtung einhergehen.
- Auf dem Objektträger ändern sich während des Kulturwachstums kritische Umweltbedingungen, wie z.B. das Nährstoffange-

bot, zeitlich und räumlich verhältnismäßig stark und erschweren die Interpretation der Ergebnisse maßgeblich.

Die Alternative zu mikroskopischen Zeitrafferaufnahmen stellten daher Wachstumsexperimente in einem kontinuierlichen Chemostat dar, aus denen in regelmäßigen Abständen Proben entnommen und diese mikroskopisch untersucht werden. Zwar lässt sich so nicht das Schicksal einzelner Zellen verfolgen, sehr wohl aber die Populationsdynamik aus den Momentanaufnahmen ableiten und in einem weiteren Schritt die gesuchten Modellparameter des Modells mathematisch ermitteln.

3.2.2 Mikroreaktionstechnische Kultivierung von *D. shibae*

3.2.2.1 Vorüberlegungen und Technik

Vor dem im letzten Kapitel beschriebenen Chemostatexperiment erschien es sinnvoll makroskopische Parameter wie die Nettowachstumsrate und Nettoproduktionsrate unter Hell- und Dunkelbedingungen durch Satzkultivierungen zu bestimmen, um optimale technische Parameter zum Betreiben eines kontinuierlichen Wachstumsexperimentes zu ermitteln. Allerdings waren die Satzkultivierungen nicht in konventionellen Schüttelkolbenkulturen durchführbar, da auch hier spätestens bei der regelmäßigen Probennahme ein unbeabsichtigter und störender Lichteintrag die Folge gewesen wäre. Alternativ sollten die Kultivierungen daher in einem automatisiertem Mikrotiterplattenverfahren durchgeführt werden. Hierzu sollte die Mikrotiterplatte nachdem sie mit Medium befüllt und einer Vorkultur inokuliert wurde, bis zum Ende des Wachstumsexperimentes in einem geeigneten Mikrotiterplattenlesegerät verbleiben und in diesem geschüttelt, passiv begast, temperiert und die einzelnen

Reaktionsgefäße durch die integrierte Messeinrichtung regelmäßig photometrisch vermessen werden.

Die Miniaturisierung von Wachstumsexperimenten hatte in der Vergangenheit eine hohe Aufmerksamkeit gewonnen (Duetz *et al.*, 2000; Kumar *et al.*, 2004; Kensy *et al.*, 2009; Rao *et al.*, 2009), da sie gegenüber herkömmlichen Wachstumsexperimenten eine Reihe von Vorteilen bietet. Diese beziehen sich z.B. auf einen niedrigeren Materialeinsatz, eine vereinfachte Erhöhung der Replikatzahl und nicht zuletzt eine vereinfachte Automatisierung der eigentlichen Experimente und Messungen. Dem stehen Einschränkungen gegenüber die in der verkleinerten Geometrie der Mikrotiterplatten begründet sind. Insbesondere der Optimierung des Sauerstoffeintrags gilt ein erhöhtes Augenmerk beim Design der Experimente und der Wahl geeigneter Mikrotiterplatten und Inkubatoren (Kensy *et al.*, 2005; Duetz, 2007; Funke *et al.*, 2009). Weiterhin stellt ein solches Verfahren neue Herausforderungen daran, relevante physiologische Parameter *in vivo* zu erfassen. So auch im hier beschriebenen Fall bezogen auf die zeitaufgelöste Quantifizierung der Produktion des BChl-*a*.

3.2.2.2 Etablierung der relativen BChl-*a* Quantifizierung

Typischerweise lässt sich BChl-*a* durch Extraktion aus den Zellen und anschließender photometrischer Messung quantifizieren (Koblízek *et al.*, 2010). Auf Versuche im Mikrotiterplattenmaßstab war ein solches *in vitro* Verfahren jedoch nicht anwendbar, da es zum einen eine deutlich größere Menge an zellulärer Substanz erfordern würde und darüber hinaus die Entnahme einer Probe den automatisierten Arbeitsfluss gestört hätte. Daher wurde nach einer Möglichkeit gesucht, die intrazelluläre Quantifizierung des BChl-*a* photometrisch *in vivo* durchzuführen.

Ein erster Ansatz dazu war die Extinktionsspektren zweier Zell-

kulturen, die unter strikt hellen und strikt dunklen Bedingungen gewachsen waren aufzunehmen. Die beiden Kulturen wurden jeweils in 50 mL Schikanekolben bis zu einer optischen Dichte von $e_{650} = 0,8$ kultiviert. Anschließend wurde von beiden Kulturen zehn Verdünnungen in SWM im Verhältnis $1/10, 2/10 \dots 1$ angefertigt und die Absorbtionsspektren in 1 nm-Schritten bei einer Wellenlänge von $\lambda = 320 \text{ nm}$ bis 1000 nm in Triplikaten aufgenommen.

Abbildung 3.14 (Oben) zeigt, dass sich die jeweiligen Spektren zunächst nur wenig unterscheiden. Hintergrund dieser Beobachtung ist, dass Extinktion auf zwei verschiedene Effekte zurückzuführen ist: Zum einen durch die Streuung der gelösten Zellen und zum anderen durch die lichtabsorbierenden Chromophore, wie der des Photosyntheseapparates. Um die BChl-*a* Konzentration also photometrisch quantifizieren zu können, war es nötig nur den reinen Absorbtionsanteil dieser konvolutierten Spektren bestimmen zu können. Hierzu wurde die Annahme getroffen, dass sich die Effekte additiv zusammensetzen und der durch Streulicht verursachte Anteil wie eine Basislinie zum reinen Absorbtionspektrum wirkt. Daher ließ sich diese Basislinie durch eine polynomische Basislinienberechnung mit einem Polynom vierten Grades (Lieber & Mahadevan-Jansen, 2003; Liland & Mevik, 2011) ermitteln (Abbildung 3.14, zweite Reihe) und von den konvolutierten Spektren abziehen (Abbildung 3.14, dritte Reihe). An dieser Stelle lassen sich bereits weitreichende Unterschiede zwischen der Zusammensetzung der Chromophoren zwischen hell und dunkel inkubierter *D. shibae* Kulturen erkennen: Während die dunkel inkubierte Kultur im gesamten sichtbare Bereich (mit Ausnahme von $\lambda = 630 \text{ nm} - 770 \text{ nm}$) eine Absorbtion aufweist, ist dies für die hell inkubierten Zellen lediglich im Bereich $\lambda = 350 \text{ nm} - 450 \text{ nm}$ erkennbar. Dies rechtfertigt auch die Wahl der Wellenlänge von 650 nm zur Bestimmung der optischen Dichte als Maß der

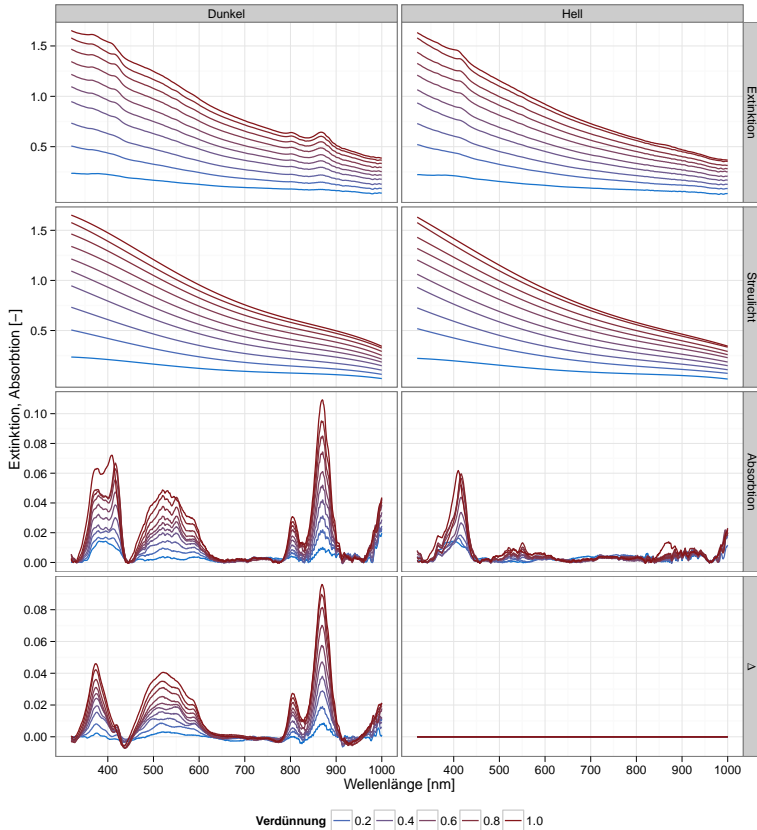


Abbildung 3.14: Dargestellt sind jeweils **oben:** Konvolvierte Extinktionsspektren (Schrittweite $\lambda = 1$ nm) einer jeweils dunkel (links) und hell (rechts) inkubierten *D. shibae* Kultur in zehn äquidistanten Verdünnungen; **zweite Reihe:** Streulichtspektren, die durch eine polynomische Basislinienanpassung der darüber stehenden Spektren errechnet wurden; **dritte Reihe:** Absorptionsspektren der Chromophore, durch Differenzbildung der Extinktionsspektren der ersten Reihe und der Streulichtspektren der zweiten Reihe; **un-**
ten: Differenz der dekonvolvierten Absorptionsspektren der dritten Reihe.

Zelldichte.

Um letztendlich die Unterschiede in der Absorbtion zwischen hell und dunkel inkubierter Kultur sehen zu können, wurden die jeweilig dekonvolutierten Absorbtionsspektren von einander abgezogen (Abbildung 3.14, unten links). Das Ergebnis zeigt das typische Absorbtionsspektrum der Chromophoren des LH2 (Robert *et al.*, 2003): Bei 374 nm ist die Soret-Bande zu erkennen, im Bereich von 450 nm bis etwa 570 nm absorbieren die Charotinoide, bei 585 nm tritt die Q_x -Bande und bei 805 nm sowie 870 nm treten die Q_y -Banden auf. Da offensichtlich im Bereich der Soret-Bande auch unter hell inkubierenden Bedingungen eine bei etwa 405 nm absorbierende Substanz detektierbar ist und sonst nur die Q_y -Banden eine verhältnismäßig scharfe, ausgeprägte und unterscheidbare Absorbtion zeigen, war zu prüfen inwiefern diese zur *in vivo* Quantifizierung geeignet sind.

Zwar liefert die hier beschriebene Dekonvolution der Spektren gute Ergebnisse, für fortlaufende Wachstumsexperimente erschien es aber unpraktikabel stets das gesamte Spektrum zu messen. Daher sollte in einem nächsten Schritt untersucht werden, ob eine lineare Interpolation der Streulichtabsorbtion an den Quantifizierungswellenlängen hinreichend genau ist. Hierzu wurden zusätzlich zur Bestimmung der Extinktionen e_{805} bei 805 nm und e_{870} bei 870 nm auch die Extinktionen e_{770} und e_{900} bestimmt, da an diesen Wellenlängen mit keiner Absorbtion mehr zu rechnen war (Abbildung 3.14, unten links). Die Ordinate s_λ der Streulichtlinie an der Wellenlänge $\lambda \in [770 \text{ nm}, 900 \text{ nm}]$ ließ sich somit durch den folgenden linearen Ansatz berechnen:

$$s_\lambda = e_{770} + \frac{e_{900} - e_{770}}{130 \text{ nm}} \cdot (\lambda - 770 \text{ nm}). \quad (3.29)$$

Die reine Absorbtion a des Chromophor an gegebener Wellenlänge

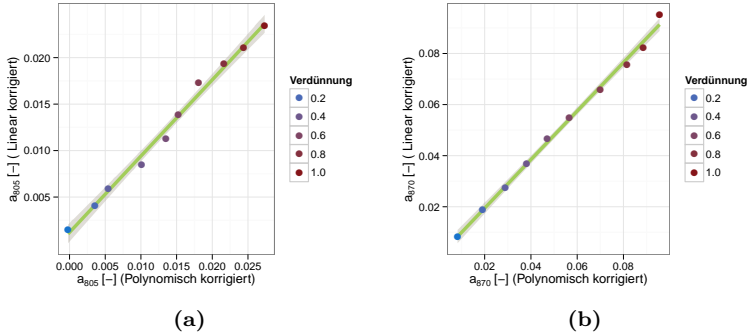


Abbildung 3.15: Dargestellt ist die lineare Regression zwischen den Werten der ermittelten Absorptionen der Q_y -Banden nach linearer gegen polynomischer Basislinienkorrektur an **a)** $\lambda = 805$ nm und **b)** $\lambda = 870$ nm. Der grau abgesetzte Bereich um die Regressionsgerade beschreibt das Konfidenzband der Regression auf einem Konfidenzniveau von 95%.

λ ließ sich entsprechend durch

$$a_\lambda = e_\lambda - s_\lambda \quad \forall \lambda \in [770, 900] \text{ nm} \quad (3.30)$$

berechnen.

Um die Qualität dieser linearen Approximation zu bestimmen, wurde eine lineare Regression der so gefundenen Werte für a_{805} und a_{870} gegen die, durch die polynomische Basislinienkorrektur gefundenen Werte an den entsprechenden Wellenlängen durchgeführt (Abbildung 3.15). Für die Regressionsgerade der Werte an $\lambda = 805$ nm konnte ein Steigung von $0,82 \pm 0,03$ und ein von 0 schwach signifikant verschiedener Achsenabschnitt von $0,0012 \pm 0,0004$ (p -Wert = 0,0235) gefunden werden. Der adjustierte, quadrierte Korrelationskoeffizient betrug $R^2 = 0,991$. Für die Regressionsgerade der Werte

an $\lambda = 870 \text{ nm}$ konnte ein Steigung von $0,95 \pm 0,02$ und ein von 0 nicht signifikant verschiedener Achsenabschnitt von $0,000 \pm 0,001$ (p -Wert = $0,726$) gefunden werden. Der adjustierte, quadrierte Korrelationskoeffizient betrug $R^2 = 0,996$. Damit konnte die lineare Basislinienapproximation insbesondere für die Wellenlänge 870 nm als hinreichend genaue Methode zur *in vivo* Quantifizierung des BChl-*a* durch Absorptionsmessung an den Q_y-Banden bestätigt werden.

3.2.2.3 Verdunstungseffekte und Schichtdickenkorrektur

Die Kultivierungsdauer von *D. shibae* Kulturen kann im Satzverfahren typischerweise über 50 h betragen. Da die Kultivierung bei 30°C durchgeführt wurden, erschien es nötig zu untersuchen, ob das Kulturmedium einer nicht zu vernachlässigenden Verdunstung ausgesetzt ist. Die Mikrotiterplatten wurden zwar mit einer hydrophoben Folie verklebt, um Kreuzkontaminationen zu vermeiden. Da diese Folie aber gasdurchlässig sein musste, konnte auch Verdunstung durch diese hindurch nicht ausgeschlossen werden.

Um diesen etwaigen Effekt zu beobachten, wurde eine Mikrotiterplatte in genau der Konfiguration befüllt und unter den selben Bedingungen wie in späteren Wachstumsexperimenten im Multi-Label Lesegerät bebrütet ohne das Medium jedoch inokuliert zu haben. Um einen Verlust an Medium zu beobachten, wurde der Umstand genutzt, dass wässrige Lösungen bei 977 nm eine deutliche Absorptionsbande zeigen, wohingegen bei 900 nm die Transmission nahezu 100% beträgt (Abbildung 3.16 (a)). Gemäß dem Lambert-Beerschen Gesetz, ist demnach die Extinktion $e_{977} - e_{900}$ proportional zur Schichtdicke h des durchstrahlten Mediums (Held, 2009). Über die Messung an diesen beiden Wellenlängen konnte daher ein möglicher Flüssigkeitsverlust über die Zeit nachvollzogen werden.

Tatsächlich erfolgte dieser auch in einem nicht zu vernachlässigen-

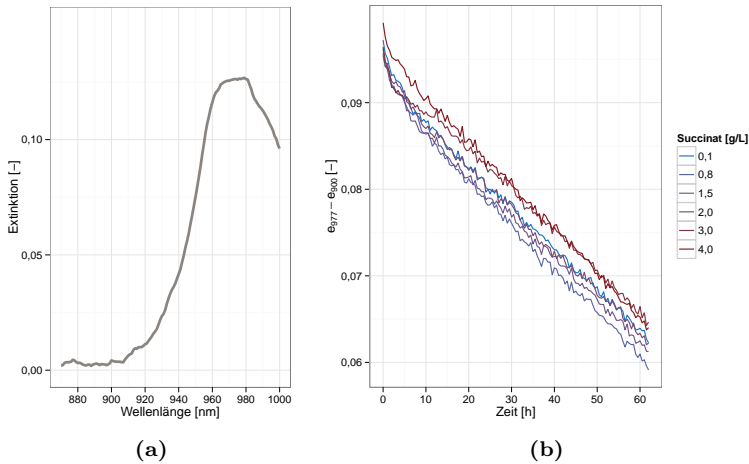


Abbildung 3.16: a) Absorptionsspektrum von SWM im NIR-Bereich. Gut zu sehen ist die Absorptionsbande von Wasser bei 977 nm. b) Zeitlicher Verlauf der Extinktionsdifferenz zwischen e_{977} und e_{900} einer ursprünglich mit 1200 μ L SWM befüllten Mikrotiterplatte. Die Endkonzentration des Succinat im Medium variierte zwischen 0,1 g/L und 4 g/L

dem Umfang (Abbildung 3.16 (b)), da die Verdunstung des Mediums über einen Zeitraum von 62 h etwa 30% mit linear fallender Tendenz betrug. Die Konsequenz aus der sich ändernden Schichtdicke ist nicht nur, dass sich die Zusammensetzung des Mediums artifiziell ändert. Insbesondere die Bestimmung der Zellkonzentration durch die Messung der optischen Dichte wird dadurch künstlich verfälscht. Daher war es auch für die folgenden Wachstumsexperimente wichtig, die entsprechenden Wellenlängen zur Schichtdickenkorrektur mitzumes-

sen. Da sich die Extinktion allerdings auch im Nahinfrarot-Bereich (NIR) durch die zunehmende Trübung im bewachsenden Medium

änderte, verbot es sich die Schichtdickenbestimmung unmittelbar in diesen Proben durchzuführen. Anstattdessen wurde sie nur in den nicht inokulierten Kavitäten der Mikrotiterplatte bestimmt, die als Blindprobe zum Hintergrundabzug ohnehin notwendig waren. Ferner wurde die Annahme getroffen, dass die Verdunstungsrate in diesen Kavitäten vergleichbar mit der in den Inokulierten war.

Um letztendlich die Schichtdicke nicht nur als Differenz zweier Absorptionen, sondern in einer Längeneinheit angeben zu können, wurde eine Kalibrierung mit verschiedenen Schichtdicken h durchgeführt. Für die hier verwendete Konfiguration aus Mikrotiterfolie, Abdeckfolie und SWM ergab sich so der Zusammenhang:

$$h = 6.49 \text{ cm} \cdot (e_{977} - e_{900}) + 0.02 \text{ cm}$$

mit einem adjustierten $R^2 = 0,994$. Unter Zuhilfenahme dieser Kalibrierung wurden alle im weiteren Verlauf gemessenen Extinktionen auf eine theoretische Schichtdicke von 1 cm umgerechnet.

3.2.2.4 Etablierende Wachstumsexperimente

Nachdem die bisher beschriebenen Vorexperimente die Frage bejahten, ob eine relative *in vivo* Quantifizierung des BChl-*a* möglich ist, sollte durch ein weiteres Experiment, diese Quantifizierung in einer wachsenden *D. shibae* Kultur im Mikrotiterplattenmaßstab durchgeführt werden. Das Wachstum fand in SWM zu Triplikaten in sechs Succinat-Konzentrationen (0.1 g/L, 0.8 g/L, 1.5 g/L, 2 g/L, 3 g/L und 4 g/L) statt. Die Kultivierung wurde über 62 h im Multi-Label Lesegerät bei 30 °C durchgeführt. Um den Sauerstofftransport ins Medium zu fördern, wurde die Mikrotiterplatte in den Phasen, in denen nicht gemessen wurde, bei 480 rpm an einem orbitalen Schüttelradius von 6 mm geschüttelt. Die relevanten Extinktionen wurden

automatisiert alle 30 min gemessen.

Das Ergebnis dieser Kultivierung ist Abbildung 3.17 zu entnehmen. Die Messwerte der einzelnen Replikate wurden nicht aufgetragen, da die relativen Fehler fast ausnahmslos deutlich unter 5% lagen und somit auch graphisch bei der gegebenen Datendichte nicht darstellbar waren. Anstatt dessen wurden nur die Mittelwerte aufgetragen. In jedem Fall verdeutlicht dieser Umstand bereits die hohe Reproduzierbarkeit des Ansatzes der Mikrotiterkultivierung.

Die Wachstumskurven haben alle prinzipiell den selben Verlauf, der sich in drei Phasen gliedert: Ein exponentieller Verlauf, der bis zu 14 h anhält, gefolgt von einem linearen Wachstum (nur in den drei höchsten Succinat-Anfangskonzentrationen), bis in der dritten Phase die Kultur abrupt stationär wird. Um zu zeigen, dass der Übergang in die Stationärphase ausschließlich durch die Verfügbarkeit der Kohlenstoffquelle determiniert ist, wurde eine lineare Regression der in jeder Kavität maximal erreichten Extinktion bei 650 nm gegen die eingesetzte Succinat-Konzentration c durchgeführt. Als Regressionsgerade konnte $e_{650}^{\max} = (0,633 \pm 0,006) L/g \cdot c + (0,11 \pm 0,01)$ mit einem adjustierten $R^2 = 0,999$ gefunden werden, womit der kausale Zusammenhang als erklärt gilt.

Der lineare Wachstumsverlauf der zweiten Phase, lässt hingegen auf eine wachstumslimitierende Komponente schließen, die ab dem Eintritt in diese Phase (etwa nach 14 h) in einer konstanten Geschwindigkeit „zugefüttert“ wird. Da die Anordnung des Experimentes allerdings nur einen Gasaustausch der Kavitäten mit der Umwelt zuließ und der Umstand, dass die zweite Phase immer ab der selben optischen Dichte von $e_{650} \approx 1$ einsetzte, ist naheliegend, dass Sauerstoff diese Limitation verursachte. Tatsächlich ist bereits bekannt, dass der Sauerstofftransfer bei Kultivierungen im Mikrotiterplattenmaßstab ein Problem gegenüber herkömmlichen Methoden, wie der

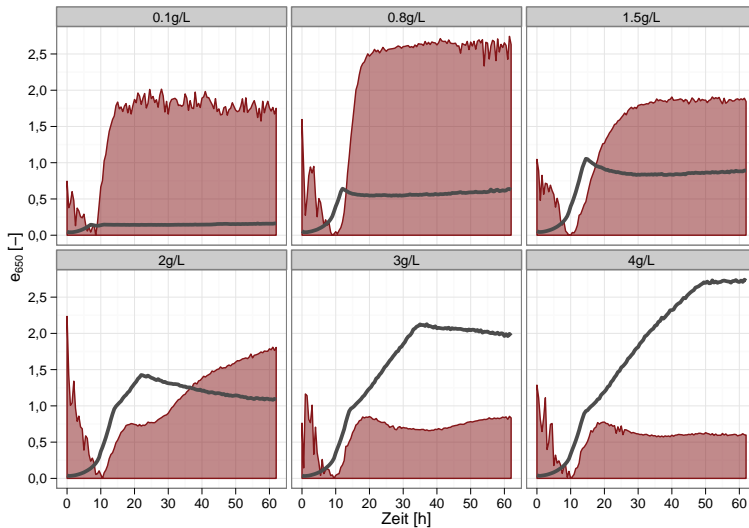


Abbildung 3.17: Dargestellt sind die Wachstumskurven (grauer Verlauf) von *D. shibae* Kulturen, deren Anzucht dunkel in einer Mikrotiterplatte bei den sechs Succinat-Konzentrationen 0.1 g/L, 0.8 g/L, 1.5 g/L, 2 g/L, 3 g/L und 4 g/L stattfand. Zusätzlich wurde die Absorption der Q_y Bande an der Wellenlänge 870 nm gemessen und jeweils auf die Extinktion bei 650 nm bezogen (roter Verlauf). Deren Werte sind nicht an der Ordinate abzulesen und wurden lediglich zur Veranschaulichung mit dem Faktor 18 in die Abbildungen skaliert.

Schüttelkolbenkultivierung darstellt. Entscheidenden Einfluss haben neben der Geometrie der Kavitäten die Schüttlerdrehzahl und der Radius. Es konnte in der Vergangenheit auch gezeigt werden (Duetz, 2007), dass eine Vergrößerung des Schüttelradius eine deutlich höhere Steigerung der Sauerstofftransferrate als eine Erhöhung der Schüttlerdrehzahl hat. Leider ließ das Instrument, das im hier betrachteten Fall eingesetzt wurde, die Variation dieser beiden Parameter nur in Grenzen zu (Thermo Scientific, 2007). Ab einem Schüttleradius oberhalb 6 mm nahm die zulässige Schüttelgeschwindigkeit so drastisch ab, dass die hier gewählte Konfiguration bereits den besten Kompromiss darstellte.

Bezüglich der intrazellulären BChl-*a* -Konzentrationen war zu beobachten, dass diese in allen Fällen zu Beginn der Kultivierung zu fallen begann. Dies ist wahrscheinlich dem Umstand geschuldet, dass die Kultur beim Animpfen über einen Zeitraum von mehreren Minuten Licht ausgesetzt war und dadurch die weitere Produktion des BChl-*a* zunächst inhibiert wurde. Durch das exponentielle Wachstum kam es dementsprechend in der Anfangsphase zu einer intrazellulären Verdünnung. In allen Kulturen begann 10 h nach Inokulierung die intrazelluläre Konzentration zu steigen. In den Fällen, in denen die Kultur dann eine Sauerstofflimitation erfuhr, wurde die Produktion zu diesem Zeitpunkt wieder gestoppt und in der Stationärphase auch nicht wieder aufgenommen. Lediglich in den Kulturen, in denen die Succinat-Anfangskonzentration 2 g/L bzw. 3 g/L betrug, ist offensichtlich auch in der Stationärphase ein weiterer Anstieg zu verzeichnen. Allerdings fällt auch die Zelldichte im gleichen Maße, so dass davon auszugehen ist, dass die in der Kultur vorhandene Gesamtabundanz von BChl-*a* konstant blieb. In den Kulturen ohne Sauerstofflimitation hingegen, stieg die Produktion auch unter nährstofflimitierten Bedingungen noch über mehrer Stunden an.

Dies kann ein Indiz dafür sein, dass unter Einwirkung von Nährstoffmangel die Produktion durch *D. shibae* weiterhin angetrieben wird, um im Fall einer späteren Lichtexposition ein erhöhtes Maß Energie sammeln und so den Nährstoffmangel kompensieren zu können.

3.2.3 Hell-Dunkel-Bedingungen im Mikroreaktionsmaßstab

3.2.3.1 Die Entwicklung eines Prototypen

Nachdem die bisher gezeigten Experimente die Machbarkeit der *in vivo* Quantifizierung von BChl-*a* in Mikrotiterkulturen bestätigten, stellte sich als nächstes die Frage, wie fluktuierende Hell-Dunkel-Phasen in diese Konfiguration zu integrieren waren. Hierbei spielten eine Reihe von Aspekten bzw. Überlegungen, die verworfen werden mussten eine Rolle:

- Die Integration einer Lichtquelle in das Instrument erschien aufgrund der baulichen Gegebenheiten nicht praktikabel.
- Ein Umsetzen der Platten in einen externen, beleuchteten Inkubator war vor dem Hintergrund der gewünschten Vollautomatisierung und einer mangelnden Robotik ebenfalls nicht praktikabel.
- Zuletzt war es nötig, die beiden Prozesse der Beleuchtung und Messung zu synchronisieren, da bei den durchgeführten photometrischen Messungen innerhalb der Messkammer auch während der Beleuchtungsphasen zwingend Dunkelheit herrschen musste.

Diese Überlegungen führten dazu, die Beleuchtung direkt in die Mikrotiterplatte zu integrieren und über die Automatisierungsschnittstelle der Instrumentsoftware anzusprechen. Zu diesem Zweck wurde

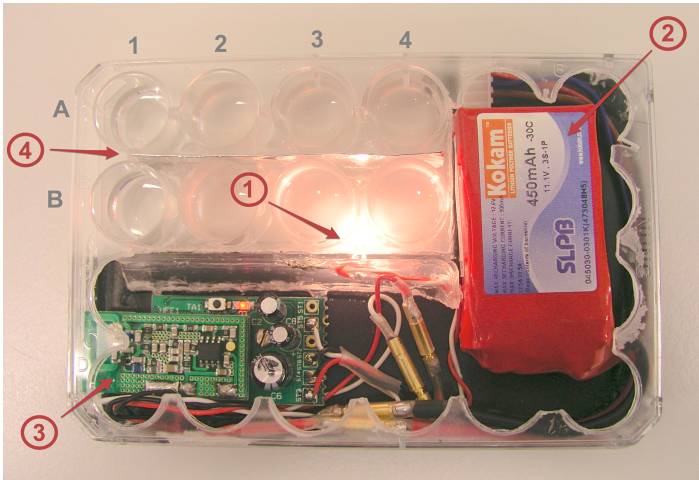


Abbildung 3.18: Prototyp einer Mikrotiterplatte mit einem integrierten, funktgesteuerten Leuchtmittel. (1) Subminiaturglühlampe zur Beleuchtung. (2) Lithium-Polymer Akkumulator zur Stromversorgung. (3) Funkschalter zum Ein- und Ausschalten des Leuchtmittels. (4) Abschattung aus Aluminiumfolie, um die Kavitäten der Reihe A möglichst dunklen Bedingungen auszusetzen. In Grau dargestellt sind die Kavitätsreihen (1-4) und Kavitätsspalten (A,B).

eine typische Mikrotiterplatte, wie sie auch im letzten Experiment verwendet wurde, durch einen Umbau mit den folgenden Funktionen versehen (Abbildung 3.18):

- Eine Glühlampe wurde als Leuchtmittel integriert. Diese schien geeignet, da sie neben der miniaturisierten Bauform anders als z.B. Leuchtdioden über ein kontinuierliches Emissionsspektrum verfügt und dieses dem des natürlichen Sonnenlichts gleicht.
- Um diese Glühlampe autark mit Strom zu versorgen, wurde ein wiederaufladbarer Lithium-Polymer Akkumulator in den

Schaltkreis integriert.

- Zur Steuerung des Leuchtmittels wurde ein 868,35 MHz Funk-schalter verbaut.
- Das Senden der Signale zum Ein- und Ausschalten des Leucht-mittels wurde durch einem externen USB-Funksender reali-siert, der die erforderliche Synchronisation vermittelte.
- Eine Abschattung aus Aluminiumfolie zwischen den Kavitä-ten und eine asymmetrische Position des Leuchtmittels soll-ten während der heller Phasen Bedingungen unterschiedlicher Lichtintensität erzwingen.

Um die synchronisierte Steuerung durchführen zu können, wur-de das Konsolenprogramm **SkanItSmartClient** in der Program-miersprache C# entwickelt. Auf die Entwicklung einer interaktiven Benutzerschnittstelle wurde verzichtet. Anstattdessen erfolgte die Steuerung über ein wohlgeformtes XML-Format, welches dem Pro-gramm als Eingabe zur Verfügung gestellt wurde und sequenziell verschiedene, in der Instrumentsoftware hinterlegte Analyse- und Inkubationsprotokolle ausführt. Eine so in XML definierte Sequenz ist im Folgenden exemplarisch dargestellt.

```

1 <?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
2 <skanItClient>
3   <clientConfig file="C:/Programme/SkanIt/ClientAI.config" />
4   <instrument name="Varioskan Flash" />
5   <user name="admin" password=" " />
6
7   <protocol step="light2-" plate="plate0" usesLight="TRUE">
8     <loop times="20">
9       <session name="DshibaeAbsorbtionLight2" />
10      <incubation name="Incubate480_6_30min_30C_short" light="off" />
11    </loop>

```

```
12     <loop times="20">
13         <session name="DshibaeAbsorbtionLight2" />
14         <incubation name="Incubate480_6_30min_30C_short" light="on"/>
15     </loop>
16     <session name="DshibaeAbsorbtionLight2" />
17 </protocol>
18 </skanitClient>
```

Zeilen 3-5 definieren allgemeine Verbindungsparameter mit der Automatisierungsschnittstelle, Zeile 7-17 die eigentliche Sequenz, die durch das Programm abgearbeitet wird. Nimmt der Parameter **uses-Light** den Wert **TRUE** an, sucht das Programm an den am Rechner verbundenen USB-Schnittstellen nach dem Funksender und bindet diesen ein. Innerhalb des **protocol**-Containers dürfen **session**-, **incubation**- und **loop**-Elemente auftreten. Die ersten beiden genannten Elemente veranlassen das Programm in der Reihenfolge in der sie auftreten, ein in der Instrumentsteuersoftware hinterlegtes Protokoll auszuführen. Der Name des Protokolls wird über den Parameter **name** definiert. Sie unterscheiden sich ausschließlich dahingehend, dass während des Ausführens eines **session**-Protokolls das Leuchtmittel immer ausgeschaltet wird bzw. bleibt, wohingegen während der Ausführung eines **incubation**-Protokolls das Leuchtmittel wahlweise eingeschaltet werden kann. Hierzu muss für den Parameter **light** der Wert **on** gewählt werden. Letztendlich dient der **loop**-Container dazu eine Sequenz von **session**- und **incubation**-Schritten in beliebig vielen Iterationen immer wieder auszuführen. Die Anzahl der Iterationen wird über den Parameter **times** gesteuert. Um komplexe Aufgaben leicht zu bewältigen, dürfen **loop**-Blöcke auch ineinander geschachtelt werden.

Der Aufruf des Programms erfolgt über die Windows-Konsole durch

```
C:\ SkanItSmartClient.exe protocoll.xml.
```


3.2.3.2 *In-vivo* BChl-*a* Messungen unter Hell-Dunkel-Bedingungen

Mit dem im letzten Abschnitt vorgestellten Prototypen wurde ein weiteres Wachstumsexperiment von *D. shibae* Kulturen unter dem Einfluss eines Dunkel-Hell-Dunkel-Zyklus durchgeführt. Die Succinat-Anfangskonzentration betrug hier in allen Fällen 2 g/L. Die Beleuchtung fand kontinuierlich im Zeitraum von 10 h bis 22 h nach Inokulierung statt und wurde nur für die kurzen Phasen der Messung deaktiviert. Da durch die Anordnung der umgebauten Mikrotiterplatte von inhomogenen Lichtbedingungen auszugehen war, wurden die Kulturen in den einzelnen Kavitäten nicht als Replikate betrachtet.

Bezüglich der Wachstumskurven (Abbildung 3.19) war zu beobachten, dass alle Kulturen der oberen Reihe, die also nur sehr indirekt Licht ausgesetzt waren, untereinander den gleichen Verlauf annahmen und dieser dem der Kultur ohne Lichtexposition (Abbildung 3.17) entsprach. In der unteren Reihe hingegen erreichten die Kultur unter direkter (Kavität B3 und B4) bzw. abgeschwächter (Kavität B2) Lichtexposition zunächst deutlich höhere Zelldichten, da zusätzlich zu der Kohlenstoffquelle die Lichtenergie zum Zellwachstum verwendet werden konnte. Allerdings brach die Zelldichte etwa nach einer Beleuchtungsdauer von 6 h (also zu dem Zeitpunkt zu dem das Substrat vollständig depletiert war) ein und fiel in etwa auf das Niveau der nicht beleuchteten Kulturen.

Die BChl-*a* Konzentrationen waren zu Beginn der Inkubation wieder in allen Fällen relativ hoch und fielen zunächst tendentiell. Anders als im vorangegangenen Experiment waren die absoluten Werte jedoch deutlich höher. Weiterhin unterlagen diese Konzentrationen größeren Schwankungen. Ursächlich dafür war das Beziehen der Absorptionen der Q_y -Bande auf die Extinktion bei 650 nm, die zu

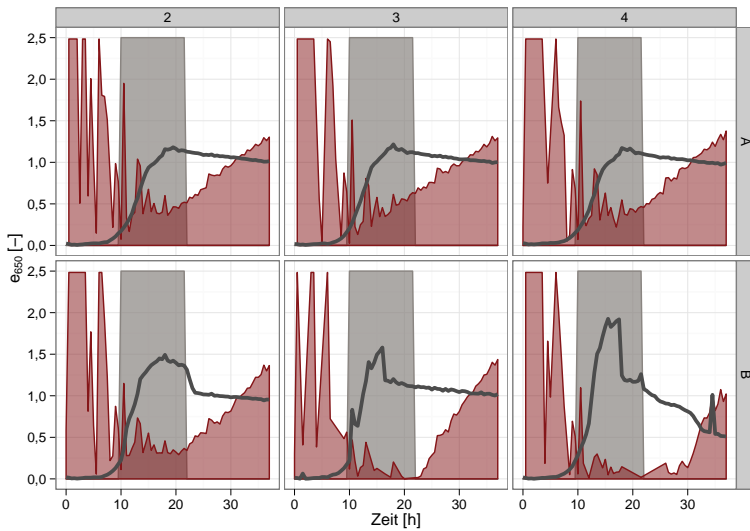


Abbildung 3.19: Dargestellt sind die Wachstumskurven (grauer Verlauf) von *D. shibae* Kulturen deren Anzucht in der umgebauten Mikrotiterplatte (Abbildung 3.18) bei der Succinat-Anfangskonzentrationen von 2 g/L stattfand. Die Beleuchtung fand kontinuierlich in der Zeit von 10 h bis 22 h nach Inokulierung statt (graue Schattierung). Die Anordnung der Diagramme entspricht der in der Mikrotiterplatte, wobei sich in den Kavitäten A1 und B2 nicht-inokulierte Blindproben befand und diese daher auch nicht dargestellt sind. Das Leuchtmittel befand sich zwischen den Kavitäten B3 und B4. Der rote Verlauf beschreibt wieder die Absorption der Q_y Bande an der Wellenlängen 870 nm, die wieder auf die Extinktion bei 650 nm bezogen ist. Diese Absorptionswerte wurden wieder mit dem selben Faktor wie in Abbildung 3.17 skaliert.

Beginn der Kultivierung auf einem niedrigen Niveau war und so kleinere Abweichungen der Zelldichte einen großen Fehler nach sich ziehen können. Ferner machte sich in diesem Fall auch das Fehlen von Replikaten bemerkbar, welche diese Fehler wahrscheinlich ausgeglichen hätten. In jedem Fall ist offensichtlich, dass der Verlust an zellulärem BChl-*a* in den direkt beleuchteten Kavitäten nach Starten der Belichtungsphase höher ausfiel und erst nach Beedingung der Beleuchtung wieder einzusetzen begann. Ebenso ist zu beobachten, dass die Kavität B2, die nur abgeschwächtem Licht ausgesetzt war, das gleiche zeitliche BChl-*a*-Profil aufweist, wie die Kulturen, die sich hinter der Abschattung befanden. Unter Umständen kann das ein Hinweis darauf sein, dass die Lichtintensität bei der Regulierung der BChl-*a* Produktion einen nicht unerheblichen Einfluss hat.

In jedem Fall haben sich diese Daten durch die erwähnten Umstände nicht für eine quantitative Auswertung angeboten. Insbesondere der Mangel an Replikaten, die unter exakt gleichen Bedingungen hätten gemessen werden können, ließ hier nur eine qualitative Auswertung zu.

3.2.3.3 Konzeptionierung eines fortgeschrittenen Prototypen

Der letzte Versuch zeigte, dass die Notwendigkeit bestand der Entwicklung einer Vorrichtung, die eine höhere Reproduzierbarkeit bei definierteren Bedingungen gewährleistet. Diese Vorrichtung unterscheidet sich durch die folgenden Erweiterungen vom ersten Prototypen (Kapitel 3.2.3.1):

- Eine höhere Anzahl zur Verfügung stehender Kavitäten.
- Eine höchstmögliche optische Isolation der Kavitäten voneinander.

- Eine individuelle Beleuchtung jeder einzelnen Kavität hinsichtlich der Beleuchtungszeiten, der Lichtintensität und letztendlich auch der Zusammensetzung des emittierten Spektrums im UV- sichtbaren und NIR-Bereiches.
- Sauerstofftransferfördernde Mittel.

Diese Vorrichtung sollte zwingend die typischen Außenmaße standardisierter Mikrotiterplatten haben (Ansi/SBS 1-4, 2004), um weiterhin mit allen im Bereich der Laborautomation gängigen Lesegeräten, Inkubatoren und Robotern kompatibel zu sein. Sowohl der geforderte Grad der Miniaturisierung, als auch der Anspruch Intensität und Emissionsspektrum individuell steuern zu können, führte bei diesen Überlegungen von Glühlampen weg hin zu einer auf Leuchtdioden (LEDs) basierenden Konstruktion (Abbildung 3.20). Da verschiedene LED-Substrate unterschiedliche Lichtspektren emittieren, lässt sich durch deren Mischen in verschiedenen Intensitäten nahezu jedes mögliche Spektrum erzeugen. Ferner liefert die Variation der Lichtintensität durch Pulsweitenmodulation, wie sie für LEDs typisch ist, über weite Intensitätsbereiche das gleiche Emissionsspektrum, wohingegen das Dimmen von Glimm- und Glühlampen mit Verschiebungen in den Lichttemperaturen einhergeht. Eine solche Konstruktion könnte also die Frage beantworten ab welchen Lichtintensitäten auf die Regulation der BChl-*a* Produktion Einfluss genommen wird und zudem welcher Lichtanteil letztendlich dafür verantwortlich ist. Desweiteren sollte die entsprechende Vorrichtung einen modularen Aufbau haben und neben einem wiederverwertbaren Gehäuse mit integrierter Elektronik den eigentlich Ort der Inkubation durch ein passendes Verbrauchsmaterial bereitstellen. Dieses ließe sich z.B. mit Schikanen ausstatten, um auch bei niedrigen Schüttleradien einen erhöhten Gasaustausch mit dem Kulturmedi-

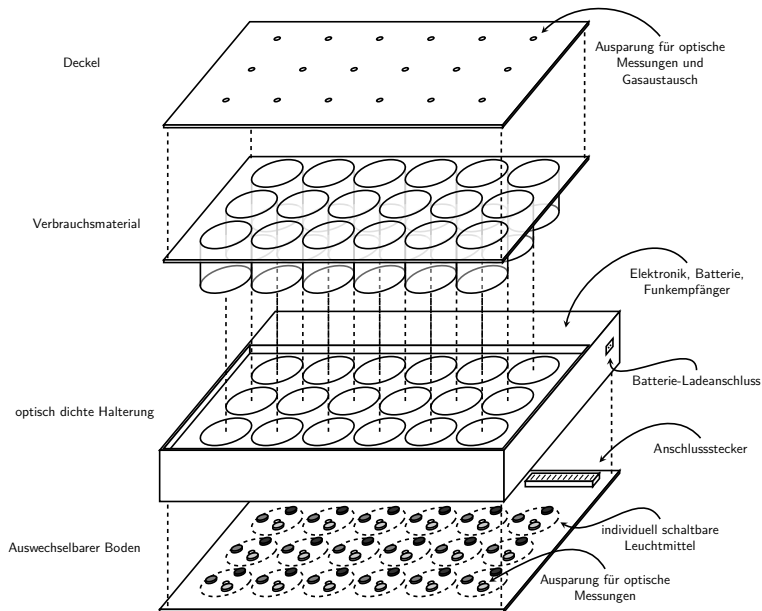


Abbildung 3.20: Konstruktionsskizze einer fortgeschrittenen Vorrichtung zur Inkubation im Mikrotitermaßstab mit integrierter Beleuchtung. Die Bodenplatte trägt pro Kavität mehrere Leuchtmittel, die bezüglich Intensität und Leuchtdauer individuell schaltbar sind. Der darüberliegende Aufbau schirmt die Kavitäten optisch voneinander ab und enthält zudem die Elektronik und Stromversorgung. In diesem Aufbau ließe sich ein optisch durchgängiges Verbrauchsmaterial einsetzen, in dem die eigentliche Inkubation stattfinden würde. Abgeschlossen wird die Konstruktion mit einem Deckel, der zwar Kreuzkontaminationen vermeidet, aber durch kleine Aussparungen (wie an der Bodenplatte) den Strahlengang für optische Messungen wie auch den Gasaustausch zulässt. Die Außendimensionen dieser Vorrichtung entsprechen dem gültigen Industrienormen für Mikrotiterplatten.

um zu fördern (Funke *et al.*, 2009).

Eine entsprechende Vorrichtung ließ sich zwar bisher nicht realisieren, das entsprechende geistige Eigentum wurde aber zwischenzeitlich am Deutschen Marken- und Patentamt angemeldet (Haddad *et al.*, 2010).

4 Zusammenfassung und Ausblick

4.1 Zusammenfassung

Die Fortschritte der modernen Mikrobiologie im Allgemeinen und neuen Hochdurchsatzmethoden im Speziellen liefern eine Fülle von Daten und Informationen. Durch geeignete Methoden und Modelle prozessiert, schaffen diese eine Basis für ein besseres Verstehen physiologischer Prozesse in mikrobiellen Populationen.

Im ersten Teil dieser Arbeit wurde die Software eSOMet entwickelt. Sie ist ein Werkzeug zur Untersuchung metabolischer Profile mit dem Ziel diese ohne weiteres Hintergrundwissen nach Ähnlichkeit hierarchisch zu clustern. Da verschiedene Cluster unterschiedliche physiologische Zustände beschreiben, lassen sich damit in einem zweiten Schritt die Metabolite identifizieren, deren Abundanzänderung für den Wechsel zwischen zwei Zuständen charakteristisch ist. Zum hochqualitativen Clustern wurde ein PCA-basierter Rauschfilter und die Methode der emergenten SOMs implementiert. Am Beispiel einer Kultivierung von *C. glutamicum* über verschiedene Wachstumsphasen und auf verschiedenen Kohlenstoffquellen, konnte die Software erfolgreich evaluiert werden. Zudem konnte gezeigt werden, dass das auf ESOMs basierende Clustern eine deutlich höhere Qualität hat, als auf klassischen Methoden wie der HCA beruhende.

Im zweiten Teil wurde eine spieltheoretisches Modell zur Untersuchung der Kulturheterogenität in mikrobiologischen Populationen entwickelt. Als Heterogenitätskriterium wurde die Anzahl von Distribusomen definiert. Dies sind Zellbestandteile, deren Abundanz einen quantitativen Einfluss auf die Teilungsrate der Mitglieder der jeweiligen Subpopulation haben und die durch Zellteilung stochastisch weiterverteilt und umweltabhängig nachproduziert werden. Zur Untersuchung des Modells wurde die Hypothese aufgestellt, dass der Photosyntheseapparat von *D. shibae* in solchen Distribusomen organisiert ist und die Produktion des BChl-*a* lichtabhängig stattfindet. Unter dieser Annahme konnte theoretisch gezeigt werden, dass sich unter fluktuierenden Umweltbedingungen oszillierend-stabile Subpopulationen ausbilden, die einen evolutionären Vorteil prägen.

Um die theoretischen Ergebnisse des zweiten Teils experimentell verifizieren zu können, wurde im letzten Teil ein Kultivierungsverfahren im Mikrotitermaßstab etabliert, welches vollautomatisiert durchgeführt werden kann, fluktuierenden Umweltbedingungen und eine hohe Replikatzahl zuließ. Hierzu wurde außerdem eine Methode zur *in vivo* Quantifizierung von BChl-*a* erfolgreich etabliert. Eine Vorrichtung, die ein solches Kultivierungsverfahren hochreproduzierbar ermöglicht wurde schlussendlich konzeptioniert.

4.2 Ausblick

Die in dieser Arbeit durchgeführten Arbeiten lassen sich als Grundlage für weitergehende Untersuchungen und Entwicklungen heranziehen.

Im Rahmen der eSOMet Softwareentwicklung wären die folgenden weitergehenden Arbeiten denkbar:

1. Wünschenswert wäre, wenn für integrative Hochdurchsatzexperimente über metabolische Profile hinausgehende molekularbiologische Daten integrierbar wären. Dies könnte sich bspw. auf Transkriptom- und Proteomprofile erstrecken, um ein tieferes Verständnis der ganzheitlichen Zusammenhänge vor einem statistischen Hintergrund zu gewinnen.
2. Bezüglich der gewählten Methoden stellen die ESOMs eine ausgezeichnete Methode für ungerichtete Clusterverfahren dar. Insbesondere in Fällen, in denen aber die verschiedenen Konditionen, denen die metabolischen Profile potentiell entspringen, bekannt sind, wäre eine zusätzliche Implementierung gerichteter Klassifizierungsverfahren wünschenswert.

Bezüglich des Replikator-Distributor-Modells bieten sich die folgenden Aspekte für weitere Arbeiten an:

1. Bisher blieben Anfangswertprobleme unerwähnt. Es bleibt zu prüfen unter welchen Anfangsbedingungen sich in einer oszillierenden Umwelt langfristig immer der selbe oder voneinander abweichende Zustände des Systems an den Übergängen einstellen.
2. Die Betrachtung sich unregelmäßig ändernder Umweltbedingungen erscheint sinnvoll, um u.U. weitere Effekte evolutionären Erfolgs beurteilen zu können. Insbesondere wenn man

auch Fitnesskosten zur Produktion der Distribusome berücksichtigen würde, könnten sich durch eine stochastische Umwelt, interessante Fitnesslandschaften und weitere spieltheoretische Betrachtungen ergeben.

3. Bisher wurde nur angenommen, dass die Wachstumsraten μ mit \mathcal{L} linear gesteigert werden (Gleichung 3.27). Vor dem Hintergrund des *Roseobacter*-Modells wäre es aber auch interessant, die Verringerung der Wachstumsraten durch stochastische Verteilungen darzustellen. Damit ließe sich z.B. nicht nur der Fitnessgewinn durch Photosynthese sondern gleichzeitig toxische Begleiterscheinungen abbilden.

Letztendlich wären im Rahmen der Mikrotiterkultivierung von *D. shibae* die folgenden Aspekte von Interesse:

1. Die Untersuchung der Lichtintensitäten und Wellenlängen, zu der die Produktion des BChl-*a* gestoppt wird, wäre von Interesse, um den Mechanismus der Regulation besser zu verstehen. Dies könnte dazu führen, die Produktion putativer Lichtsensoren genetisch zu blockieren und die Fitness einer solchen Mutante mit der des Wildtypes zu vergleichen.
2. Bezogen auf den Prozess der Mikrotiterplattenkultivierung wäre es in jedem Fall wünschenswert, den Sauerstofftransfer in das Medium z.B. durch strombrechende Einbauten in den Kavitäten weiter zu optimieren.
3. Für beide vorangehenden Fragestellungen wäre es zwingend nötig einen weiteren Prototypen, wie er in Kapitel 3.2.3.2 beschrieben wurde auch konkret umzusetzen.

Literaturverzeichnis

- (2004). *ANSI/SBS 1-2004 through ANSI/SBS 4-2004: Microplates - Footprint Dimensions, Height Dimensions, Bottom Outside Flange Dimensions, Microplates Well Positions*. Society for Bio-molecular Science.
- Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T. & Ikemura, T. (2003). Informatics for unveiling hidden genome signatures. *Genome Res*, **13** (4), 693–702.
- Allen, J. F. & Martin, W. (2007). Evolutionary biology: out of thin air. *Nature*, **445** (7128), 610–612.
- Allgaier, M., Uphoff, H., Felske, A. & Wagner-Döbler, I. (2003). Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. *Appl Environ Microbiol*, **69** (9), 5051–5059.
- Arber, W. & Linn, S. (1969). DNA modification and restriction. *Annu Rev Biochem*, **38**, 467–500.
- Askenazi, M., Driggers, E. M., Holtzman, D. A., Norman, T. C., Iverson, S., Zimmer, D. P., Boers, M.-E., Blomquist, P. R., Martinez, E. J., Monreal, A. W., Feibelman, T. P., Mayorga, M. E., Maxon, M. E., Sykes, K., Tobin, J. V., Cordero, E., Salama, S. R., Trueheart, J., Royer, J. C. & Madden, K. T. (2003). Integrating transcriptional and metabolite profiles to direct the engineering

- of lovastatin-producing fungal strains. *Nat Biotechnol*, **21** (2), 150–156.
- Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. *Nat Rev Microbiol*, **4** (8), 577–587.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2011). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. 13. überarb. Auflage, Springer Berlin Heidelberg.
- Backhaus, K., Erichson, B. & Weiber, R. (2010). *Fortgeschrittene Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer Berlin Heidelberg.
- Bellinger, G., Castro, D. & Mills, A. (2004). Webseite: <http://systems-thinking.org/dikw/dikw.htm#gjd>.
- Bergmiller, T., Peña-Miller, R., Boehm, A. & Ackermann, M. (2011). Single-cell time-lapse analysis of depletion of the universally conserved essential protein YgjD. *BMC Microbiol*, **11**, 118.
- Biebl, H., Allgaier, M., Tindall, B. J., Koblizek, M., Lünsdorf, H., Pukall, R. & Wagner-Döbler, I. (2005). *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int J Syst Evol Microbiol*, **55** (Pt 3), 1089–1096.
- Biebl, H. & Wagner-Döbler, I. (2006). Growth and bacteriochlorophyll a formation in taxonomically diverse aerobic anoxygenic phototrophic bacteria in chemostat culture: influence of light regimen and starvation. *Process Biochemistry*, **41** (10), 2153–2159.
- Bishop, D. T. & Cannings, C. (1978). A generalized war of attrition. *J Theor Biol*, **70** (1), 85–124.

- Borland, C., Cogdell, R., Land, E. & Truscott, T. (1989). Bacteriochlorophyll a triplet state and its interactions with bacterial carotenoids and oxygen. *Journal of Photochemistry and Photobiology B: Biology*, **3** (2), 237–245.
- Brehm-Stecher, B. F. & Johnson, E. A. (2004). Single-cell microbiology: tools, technologies, and applications. *Microbiol Mol Biol Rev*, **68** (3), 538–559.
- Bruhn, J. B., Gram, L. & Belas, R. (2007). Production of antibacterial compounds and biofilm formation by *Roseobacter* species are influenced by culture conditions. *Appl Environ Microbiol*, **73** (2), 442–450.
- Buchan, A., González, J. M. & Moran, M. A. (2005). Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol*, **71** (10), 5665–5677.
- Béjà, O., Suzuki, M. T., Heidelberg, J. F., Nelson, W. C., Preston, C. M., Hamada, T., Eisen, J. A., Fraser, C. M. & DeLong, E. F. (2002). Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature*, **415** (6872), 630–633.
- Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P. & Karp, P. D. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*, **40** (Database issue), D742–D753.
- Caswell, H. (2001). *Matrix Population Models*. 2nd edition, Sinauer Associates, Inc. Publishers.

- Chen, P. E., Willner, K. M., Butani, A., Dorsey, S., George, M., Stewart, A., Lentz, S. M., Cook, C. E., Akmal, A., Price, L. B., Keim, P. S., Mateczun, A., Brahmbhatt, T. N., Bishop-Lilly, K. A., Zwick, M. E., Read, T. D. & Sozhamannan, S. (2010). Rapid identification of genetic modifications in *Bacillus anthracis* using whole genome draft sequences generated by 454 pyrosequencing. *PLoS One*, **5** (8), e12397.
- Cohen-Bazire, G., Pfennig, N. & Kunisawa, R. (1964). The fine structure of green bacteria. *J Cell Biol*, **22**, 207–225.
- Cottrell, M. T., Mannino, A. & Kirchman, D. L. (2006). Aerobic anoxygenic phototrophic bacteria in the mid-atlantic bight and the north pacific gyre. *Appl Environ Microbiol*, **72** (1), 557–564.
- Cressman, R. (2003). *Evolutionary dynamics and extensive form games*. The MIT Press.
- Danna, K. & Nathans, D. (1971). Specific cleavage of simian virus 40 dna by restriction endonuclease of hemophilus influenzae. *Proc Natl Acad Sci USA*, **68** (12), 2913–2917.
- Delbrück, M. (1945). The burst size distribution in the growth of bacterial viruses (bacteriophages). *J Bacteriol*, **50**, 131–135.
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J Royal Stat Soc, Series B*, **39** (1), 1–38.
- Duetz, W. A. (2007). Microtiter plates as mini-bioreactors: miniaturization of fermentation methods. *Trends Microbiol*, **15** (10), 469–475.

- Duetz, W. A., Rüedi, L., Hermann, R., O'Connor, K., Büchs, J. & Witholt, B. (2000). Methods for intense aeration, growth, storage, and replication of bacterial strains in microtiter plates. *Appl Environ Microbiol*, **66** (6), 2641–2646.
- Eggeling, L. & Bott, M. (2005). *Handbook of Corynebacterium glutamicum*. CRC Press, Boca Raton, USA.
- Eigen, M. (1971). Molecular self-organization and the early stages of evolution. *Q Rev Biophys*, **4** (2), 149–212.
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*, **95** (25), 14863–14868.
- Eldar, A. & Elowitz, M. (2005). Systems biology: deviations in mating. *Nature*, **437** (7059), 631–632.
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, **297** (5584), 1183–1186.
- Fiehn, O. (2001). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp Funct Genomics*, **2** (3), 155–168.
- Fiehn, O. (2002). Metabolomics—the link between genotypes and phenotypes. *Plant Mol Biol*, **48** (1-2), 155–171.
- Fiehn, O. & Kind, T. (2006). *Metabolomics: Methods and Protocols*. Kapitel Metabolite Profiling in Blood Plasma, pp. 3 – 17. Springer.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty,

- B. A. & Merrick, J. M. (1995). Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science*, , **269** (5223), 496–512.
- Frimmersdorf, E. (2010). *Metabolomanalyse von Corynebacterium glutamicum und Pseudomonas aeruginosa unter verschiedenen Wachstumsbedingungen*. Doktorarbeit, Technische Universität Braunschweig.
- Funke, M., Diederichs, S., Kensy, F., Müller, C. & Büchs, J. (2009). The baffled microtiter plate: increased oxygen transfer and improved online monitoring in small scale fermentations. *Biotechnol Bioeng*, , **103** (6), 1118–1128.
- Fürch, T., Preusse, M., Tomasch, J., Zech, H., Wagner-Döbler, I., Rabus, R. & Wittmann, C. (2009). Metabolic fluxes in the central carbon metabolism of *Dinoroseobacter shibae* and *Phaeobacter gallaeciensis*, two members of the marine *Roseobacter* clade. *BMC Microbiol*, , **9**, 209.
- Gamma, E., Helm, R. & Johnson, R. (2001). *Entwurfsmuster – Elemente wiederverwendbarer objektorientierter Software*. 2. aufl., Addison-Wesley.
- Garfinkel, D., Garfinkel, L., Pring, M., Green, S. B. & Chance, B. (1970). Computer applications to biochemical kinetics. *Annu Rev Biochem*, , **39**, 473–498.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, , **81** (25), 2340–2361.
- Goto, S., Okuno, Y., Hattori, M., Nishioka, T. & Kanehisa, M.

- (2002). LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res*, **30** (1), 402–404.
- Grima, R. & Schnell, S. (2008). Modelling reaction kinetics inside cells. *Essays Biochem*, **45**, 41–56.
- Grinstead, C. M. & Snell, J. L. (1997). *Introduction to Probability* Kapitel 11 Markov Chains, pp. 405 – 470. American Mathematical Society.
- Grosu, P., Townsend, J. P., Hartl, D. L. & Cavalieri, D. (2002). Pathway processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res*, **12** (7), 1121–1126.
- Haddad, I., Hiller, K., Frimmersdorf, E., Benkert, B., Schomburg, D. & Jahn, D. (2009). An emergent self-organizing map based analysis pipeline for comparative metabolome studies. *In Silico Biol*, **9** (4), 163–178.
- Haddad, I., Münch, R. & Jahn, D. (2010). Patentanmeldung beim deutschen Patent und Markenamt: Vorrichtung und System zur Durchführung von zeitaufgelösten Untersuchungen an Zellproben. (DE: 10 2010 048 443.1).
- Heinrich, R., Rapoport, S. M. & Rapoport, T. A. (1977). Metabolic regulation and mathematical models. *Prog Biophys Mol Biol*, **32** (1), 1–82.
- Held, P. (2009). Nucleic acid quantitation using biotek’s scanning microplate spectrophotometer. Technical report, Applications Department, BioTek Instruments, Inc., USA.
- Herrero, J., Valencia, A. & Dopazo, J. (2001). A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics*, **17** (2), 126–136.

- Hiller, K. (2006). *Systems biology tools for the analysis of cellular processes*. Doktorarbeit,, Technische Universität Braunschweig.
- Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T. & Saito, K. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, , **101** (27), 10205–10210.
- Hofbauer, J. & Sigmund, K. (2003). Evolutionary game dynamics. *Bull. Amer. Math. Soc.*, **40**, 479–519.
- Ingham, C. J., Beerthuyzen, M. & van Hylckama Vlieg, J. (2008). Population heterogeneity of lactobacillus plantarum wcfsl microcolonies in response to and recovery from acid stress. *Appl Environ Microbiol*, , **74** (24), 7750–7758.
- Jiao, N., Zhang, F. & Hong, N. (2010). Significant roles of bacteriochlorophyll a supplemental to chlorophyll a in the ocean. *ISME J*, , **4** (4), 595–597.
- Jonsson, P., Gullberg, J., Nordström, A., Kusano, M., Kowalczyk, M., Sjöström, M. & Moritz, T. (2004). A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem*, , **76** (6), 1738–1745.
- Junker, B. H., Klukas, C. & Schreiber, F. (2006). Vanted: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, , **7**, 109.
- Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H. & Ikemura, T. (2001). Analysis of codon usage diversity

- of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene*, , **276** (1-2), 89–99.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T. & Yamanishi, Y. (2008). Kegg for linking genomes to life and the environment. *Nucleic Acids Res*, , **36** (Database issue), D480–D484.
- Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, , **28** (1), 27–30.
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. & Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*, , **34** (Database issue), D354–D357.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. (2012). Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*, , **40** (Database issue), D109–D114.
- Kell, D. B. & Knowles, J. D. (2006). *Systems Modelling in Cellular Modeling: from concepts to nuts and bolts*, Kapitel 1 The Role of Modeling in Systems Biology, pp. 3 – 18. The MIT Press.
- Kensy, F., Engelbrecht, C. & Büchs, J. (2009). Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in *Escherichia coli* and *Hansenula polymorpha* fermentations. *Microb Cell Fact*, , **8**, 68.

- Kensy, F., Zimmermann, H. F., Knabben, I., Anderlei, T., Trauthwein, H., Dingerdissen, U. & Büchs, J. (2005). Oxygen transfer phenomena in 48-well microtiter plates: determination by optical monitoring of sulfite oxidation and verification by real-time measurement during microbial growth. *Biotechnol Bioeng*, , **89** (6), 698–708.
- Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñoz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P. & Karp, P. D. (2011). EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res*, , **39** (Database issue), D583–D590.
- Kitano, H. (2002). Systems biology: a brief overview. *Science*, , **295** (5560), 1662–1664.
- Koblížek, M., Mlcousková, J., Kolber, Z. & Kopecký, J. (2010). On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Arch Microbiol*, , **192** (1), 41–49.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, , **43**, 59–69.
- Kohonen, T. (2001). *Self-Organizing Maps*. 3rd edition, Springer-Verlag, Berlin.
- Kolber, Z. S., Plumley, F. G., Lang, A. S., Beatty, J. T., Blankenship, R. E., VanDover, C. L., Vetriani, C., Koblizek, M., Rathgeber, C. & Falkowski, P. G. (2001). Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science*, **292** (5526), 2492–2495.

- Kolber, Z. S., Van Dover, C. L., Niederman, R. A. & Falkowski, P. G. (2000). Bacterial photosynthesis in surface waters of the open ocean. *Nature*, **407** (6801), 177–179.
- Kopka, J., Fernie, A., Weckwerth, W., Gibon, Y. & Stitt, M. (2004). Metabolite profiling in plant biology: platforms and destinations. *Genome Biol*, **5** (6), 109.
- Kouskoumvekaki, I., Yang, Z., Jónsdóttir, S. O., Olsson, L. & Pagniotou, G. (2008). Identification of biomarkers for genotyping *Aspergilli* using non-linear methods for clustering and classification. *BMC Bioinformatics*, **9**, 59.
- Kumar, S., Wittmann, C. & Heinzle, E. (2004). Minibioreactors. *Biotechnol Lett*, **26** (1), 1–10.
- Lieber, C. A. & Mahadevan-Jansen, A. (2003). Automated method for subtraction of fluorescence from biological raman spectra. *Appl Spectrosc*, **57** (11), 1363–1367.
- Liland, K. H. & Mevik, B.-H. (2011). *baseline: Baseline Correction of Spectra*. R package version 1.0-1.
- Locke, J. C. W. & Elowitz, M. B. (2009). Using movies to analyse gene circuit dynamics in single cells. *Nat Rev Microbiol*, **7** (5), 383–392.
- Lotka, A. (1934). *Théorie analytique des associations biologiques. Première partie. Principes*.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, (Cam, L. M. L. & Neyman, J., eds), vol. 1, pp. 281–297, University of California Press.

- McAdams, H. H. & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA*, **94** (3), 814–819.
- McAdams, H. H. & Arkin, A. (1999). It’s a noisy business! genetic regulation at the nanomolar scale. *Trends Genet*, **15** (2), 65–69.
- Meinicke, P., Lingner, T., Kaefer, A., Feussner, K., Göbel, C., Feussner, I., Karlovsky, P. & Morgenstern, B. (2008). Metabolite-based clustering and visualization of mass spectrometry data using one-dimensional self-organizing maps. *Algorithms Mol Biol*, **3**, 9.
- Mlecnik, B., Scheideler, M., Hackl, H., Hartler, J., Sanchez-Cabo, F. & Trajanoski, Z. (2005). Pathwayexplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res*, **33** (Web Server issue), W633–W637.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986). Specific enzymatic amplification of dna in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*, **51 Pt 1**, 263–273.
- Murray, J. D. (2002). *Mathematical Biology: I. An Introduction*. 3rd ed., corr. 2nd printing (2. june 2008), Springer New York.
- Nash, J. F. (1950). Equilibrium points in n-person games. *Proc Natl Acad Sci USA*, **36** (1), 48–49.
- Novais, R. C. & Thorstenson, Y. R. (2011). The evolution of pyrosequencing® for microbiology: from genes to genomes. *J Microbiol Methods*, **86** (1), 1–7.
- Novak, M. (2006). *Evolutionary dynamics: exploring the equations of life*. Harvard University Press.

- Onsager, L. (1931). Reciprocal relations in irreversible processes. *Physical Review*, **37** (4), 405–426.
- Oz, A., Sabehi, G., Koblízek, M., Massana, R. & Bédj, O. (2005). *roseobacter*-like bacteria in red and mediterranean sea aerobic an-oxygenic photosynthetic populations. *Appl Environ Microbiol*, **71** (1), 344–353.
- Pan, D., Sun, N., Cheung, K.-H., Guan, Z., Ma, L., Holford, M., Deng, X. & Zhao, H. (2003). PathMAPA: a tool for displaying gene expression and performing statistical tests on metabolic pathways at multiple levels for *Arabidopsis*. *BMC Bioinformatics*, **4**, 56.
- Piekarski, T. (2011). *Etablierung methodischer Grundlagen und in-itiiale Untersuchungen zur physiologischen Charakterisierung von Bakterien der Roseobacter-Gruppe*. Doktorarbeit, Technische Uni-versität Braunschweig.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- Psencík, J., Ikonen, T. P., Laurinmäki, P., Merckel, M. C., Butcher, S. J., Serimaa, R. E. & Tuma, R. (2004). Lamellar organization of pigments in chlorosomes, the light harvesting complexes of green photosynthetic bacteria. *Biophys J*, **87** (2), 1165–1172.
- Rao, G., Moreira, A. & Brorson, K. (2009). Disposable bioproc-essing: the future has arrived. *Biotechnol Bioeng*, **102** (2), 348–356.
- Robert, B., Cogdell, R. J. & van Grondell, R. (2003). The light-harvesting system of purple bacteria. In *Light-Harvesting Anten-nas in Photosynthesis*. Kluwer Academic Publishers.

- Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L. & Fernie, A. (2001). Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell*, **13** (1), 11–29.
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, 53–65.
- Sambrook, J. & Russell, D. (2001). *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press.
- Sandholm, W. H. (2010). *Population games and evolutionary dynamics*. The MIT Press.
- Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J. & Schomburg, D. (2011). BRENDA, the enzyme information system in 2011. *Nucleic Acids Res*, **39** (Database issue), D670–D676.
- Schneider, M. V. & Orchard, S. (2011). Omics technologies, data and bioinformatics principles. *Methods Mol Biol*, **719**, 3–30.
- Schrödinger, E. (1943). *What is life?* Cambridge University Press.
- Schuster, P. & Sigmund, K. (1983). Replicator dynamics. *J theor Biol*, **100**, 533–538.
- Shiba, T. (1991). *roseobacter litoralis* gen. nov., sp. nov., and *roseobacter denitrificans* sp. nov., aerobic pink-pigmented bacteria which contain bacteriochlorophyll a. *Syst Appl Microbiol*, **14**, 140–145.
- Smith, J. M. (1974). The theory of games and the evolution of animal conflicts. *J Theor Biol*, **47** (1), 209–221.

- Smith, J. M. (1976). Evolution and the theory of games. *Am Sci*, **64** (1), 41–45.
- Soetaert, K., Petzoldt, T. & Setzer, R. W. (2010). Solving differential equations in r: package desolve. *Journal of Statistical Software*, **33** (9), 1–25.
- Spudich, J. L. & Koshland, Jr, D. (1976). Non-genetic individuality: chance in the single cell. *Nature*, **262** (5568), 467–471.
- Szabó, G. & Fáth, G. (2007). Evolutionary games on graphs. *Physics Reports*, **446** (4-6), 97 – 216.
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. & Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA*, **96** (6), 2907–2912.
- ThermoScientific (2007). *VarsioskanFlash User Manual* rev 1.1.
- Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y. & Stitt, M. (2004). MAP-MAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J*, **37** (6), 914–939.
- Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., Umezawa, T., Misawa, N., Saito, K. & Shibata, D. (2005). KaPPA-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol*, **138** (3), 1289–1300.
- Tomasch, J., Gohl, R., Bunk, B., Diez, M. S. & Wagner-Döbler, I. (2011). Transcriptional response of the photoheterotrophic marine

- bacterium *Dinoroseobacter shibae* to changing light regimes. *ISME J*, **2011**, 1–12.
- Ultsch, A. (1999). *Kohonen Maps* Kapitel Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series, pp. 34 – 47. Elsevier.
- Ultsch, A. (2003). U*-matrix: a tool to visualize clusters in high dimensional data. Technical Report 36, University of Marburg, Department of Computer Science.
- van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, , **7**, 142.
- Veening, J.-W., Smits, W. K. & Kuipers, O. P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol*, , **62**, 193–210.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. D., Dunn,

P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferreira, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely,

- D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. (2001). The sequence of the human genome. *Science*, **291** (5507), 1304–1351.
- Villas-Bôas, S. G., Moxley, J. F., Akesson, M., Stephanopoulos, G. & Nielsen, J. (2005). High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem J*, **388** (Pt 2), 669–677.
- Villas-Bôas, S. G., Nielsen, J., Smedsgaard, J., Hansen, M. E. & Roessner-Tunali, U. (2007). *Metabolome Analysis: An Introduction*. Wiley-Interscience.
- Wagner-Döbler, I., Ballhausen, B., Berger, M., Brinkhoff, T., Buchholz, I., Bunk, B., Cypionka, H., Daniel, R., Drepper, T., Gerdts, G., Hahnke, S., Han, C., Jahn, D., Kalhoefer, D., Kiss, H., Klenk, H.-P., Kyrpides, N., Liebl, W., Liesegang, H., Meincke, L., Pati, A., Petersen, J., Piekarski, T., Pommerenke, C., Pradella, S., Pukall, R., Rabus, R., Stackebrandt, E., Thole, S., Thompson, L., Tielen, P., Tomasch, J., von Jan, M., Wanphrut, N., Wichels, A., Zech, H. & Simon, M. (2010). The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a hitchhiker's guide to life in the sea. *ISME J*, **4** (1), 61–77.

- Wagner-Döbler, I. & Biebl, H. (2006). Environmental biology of the marine *Roseobacter* lineage. *Annu Rev Microbiol*, **60**, 255–280.
- Wang, D. & Bodovitz, S. (2010). Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol*, **28** (6), 281–290.
- Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171** (4356), 737–738.
- Wendisch, V. F., de Graaf, A. A., Sahm, H. & Eikmanns, B. J. (2000). Quantitative determination of metabolic fluxes during coutilization of two carbon sources: comparative analyses with *Corynebacterium glutamicum* during growth on acetate and/or glucose. *J Bacteriol*, **182** (11), 3088–3096.
- Westerhoff, H. V. & Palsson, B. O. (2004). The evolution of molecular biology into systems biology. *Nat Biotechnol*, **22** (10), 1249–1252.
- Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.
- Wurtele, E., Li, J., Diao, L., Zhang, H., Foster, C., Fatland, B., Dickerson, J., Brown, A., Cox, Z., Cook, D., Lee, E.-K. & Hofmann, H. (2003). MetNet: Software to build and model the biogenetic lattice of *Arabidopsis*. *Comp Funct Genom*, **4**, 239–245.
- Young, J. W., Locke, J. C. W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P. S., Mjolsness, E. & Elowitz, M. B. (2012). Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc*, **7** (1), 80–88.
- Yurkov, V. V. & Beatty, J. T. (1998). Aerobic anoxygenic phototrophic bacteria. *Microbiol Mol Biol Rev*, **62** (3), 695–724.

- Yurkov, V. V. & Gernerden, H. V. (1993). Abundance and salt tolerance of obligately aerobic, phototrophic bacteria in a marine microbial mat. *Netherlands Journal of Sea Research*, **31** (1), 57 – 62.

Danksagung



Ich war am Institut für Mikrobiologie fast durchgängig seit November 2002 in verschiedenen Funktionen und mit verschiedenen Aufgaben tätig. Was ich dort fand waren Freunde, Herausforderungen, Anerkennung und ein solides Fundament meiner persönlichen wie auch wissenschaftlichen Entwicklung. Für die phantastische Atmosphäre die dort herrschte, die ausgezeichneten Rahmenbedingungen die ich vorfand; vor allem für die Inspiration, die uns alle auf eine so besondere Weise geformt hat, gilt mein größter Dank meinem Mentor Dieter Jahn.

Max Schobert war derjenige, der jederzeit gute Laune verströmte und dem oft hektischen Alltag stets mit Humor begegnete. Alleine dafür gilt ihm von ganzem Herzen größter Dank. Er hat darüber hinaus meine Doktorarbeit über die gesamte Zeit begleitet, immer wieder wertvolle Impulse gegeben, mit seiner außerordentlich fachübergreifenden Kompetenz die ein oder andere Sackgasse aufgelöst

und sich nicht zuletzt zur Übernahme des Koreferates erklärt.

Christoph Wittmann danke ich für die prompte und freundliche Übernahme des Vorsitzes meiner Prüfungskommission.

Richard Münch danke ich für über drei Jahre intensiver und hervorragender Betreuung. Sein Instinkt leitete mich stets durch die Fahrwasser der Bioinformatik. Aus kurzen Ideen formte er in aufregender Geschwindigkeit spannende Fragestellungen, die nie anspruchslos waren. Gleichzeitig hielt er mir jederzeit wenn nötig den Rücken frei und bot mir alle Freiheiten und Möglichkeiten, die man sich als Doktorand nur wünschen konnte.

Bei Ida Retter, Andreas Grote, Johannes Klein, Boyke Bunk, Ilona Biegler, Louisa Roselius, Svitlana Yarmolinetz und Claudia Pommerenke bedanke ich mich für eine tolle, freundschaftliche Atmosphäre. Mit Euch immer wieder an einem Strang gezogen, Praktika konzipiert, Server administriert und oft bis in die Morgenstunden gefeiert zu haben, machte das gemeinsame Arbeiten zu einem mir immer in Erinnerung bleibenden Erlebnis.

Karsten Hiller gilt mein Dank für die so erfolgreiche Zusammenarbeit im eSOMet-Projekt. Mir die Möglichkeit zu geben auf seinen Arbeiten aufzubauen und sie fortzuführen war ein Zeichen großer Anerkennung. Ferner bedanke ich mich bei Jörn Pons-Kühnemann für eine äußerst umfassende Diskussion verschiedener multivariater Verfahren und Dietmar Schomburg sowie Eliane Frimmersdorf für das zur Verfügungstellen der Daten aus den *C. glutamicum* Versuchen.

Der gesamten Roseobacter-Gruppe gilt mein außerordentlicher Dank für die so freundliche und ermutigende Wiedereingliederung ins Labor: Allen voran Petra Tielen, Sarah Finke, Sebastian Laaß und Tanja Piekarski. Im Rahmen des Roseobacter-Projektes erfuhr ich zudem durch aufschlussreiche Diskussionen und praktische Hilfe viel

Unterstützung von Personen anderer Institute und Einrichtungen, so dass Jürgen Tomasch (Diskussionen zur Biochemie von *D. shibae*), Manfred Rohde (elektronenmikroskopische Aufnahmen), Jomo Walla (Diskussion zur Chlorophyllfluoreszenz) und Meinhard Schilling (Diskussion über Leuchtmittel und elektronische Schaltungen) nicht unerwähnt bleiben dürfen.

Ich bedanke mich auch bei Jörg Saathoff von der Wissenschaftstransferstelle der Technischen Universität Braunschweig und Rolf Kröncke von der Kanzlei Gramm, Lins & Partner für das Begleiten durch den im Rahmen dieser Arbeit angestoßenen Patentierungsprozess und die ein oder andere Lektion durch das europäische Patentwesen.

Im Laufe der Zeit habe ich am Institut langjährige Freundschaften pflegen und neue knüpfen können. Ganz besonders Tanja Piekarski, Rebekka Biedendieck, Simon Stammen, Ines Gruner, Anika March und Melanie Burghartz und vielen die bereits erwähnt wurden, möchte ich für eine unvergessliche Zeit während und nach der Arbeit danken.

Meinen neuen Kollegen und Freunden bei der Metabolomic Discoveries GmbH bin ich für die viele Geduld in den eineinhalb Jahren des Zusammenschreibens dankbar: Trotz der enormen Aufgaben, die wir gemeinsam bewältigen, habe ich regelmäßig die Möglichkeit und alle Unterstützung erhalten dieses Werk zu Ende zu bringen.

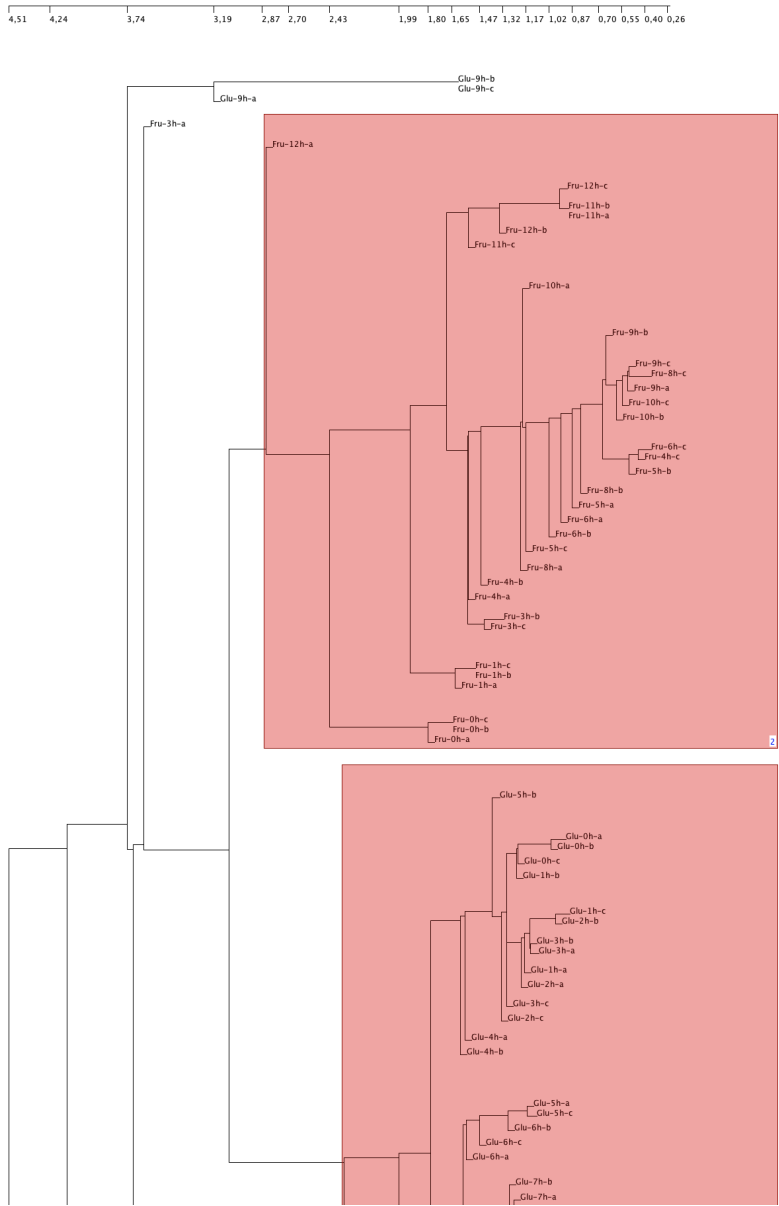
Meinen Eltern, meiner Schwester und meiner Oma danke ich für alles was sie mir mit auf den Weg gegeben haben: Meine Leidenschaft für die Naturwissenschaften, meine humanistische Bildung, den Glauben an die eigene Kreativität, nie den interdisziplinären Gedanken verloren zu haben und auch in Durststrecken nicht aufgegeben zu haben.

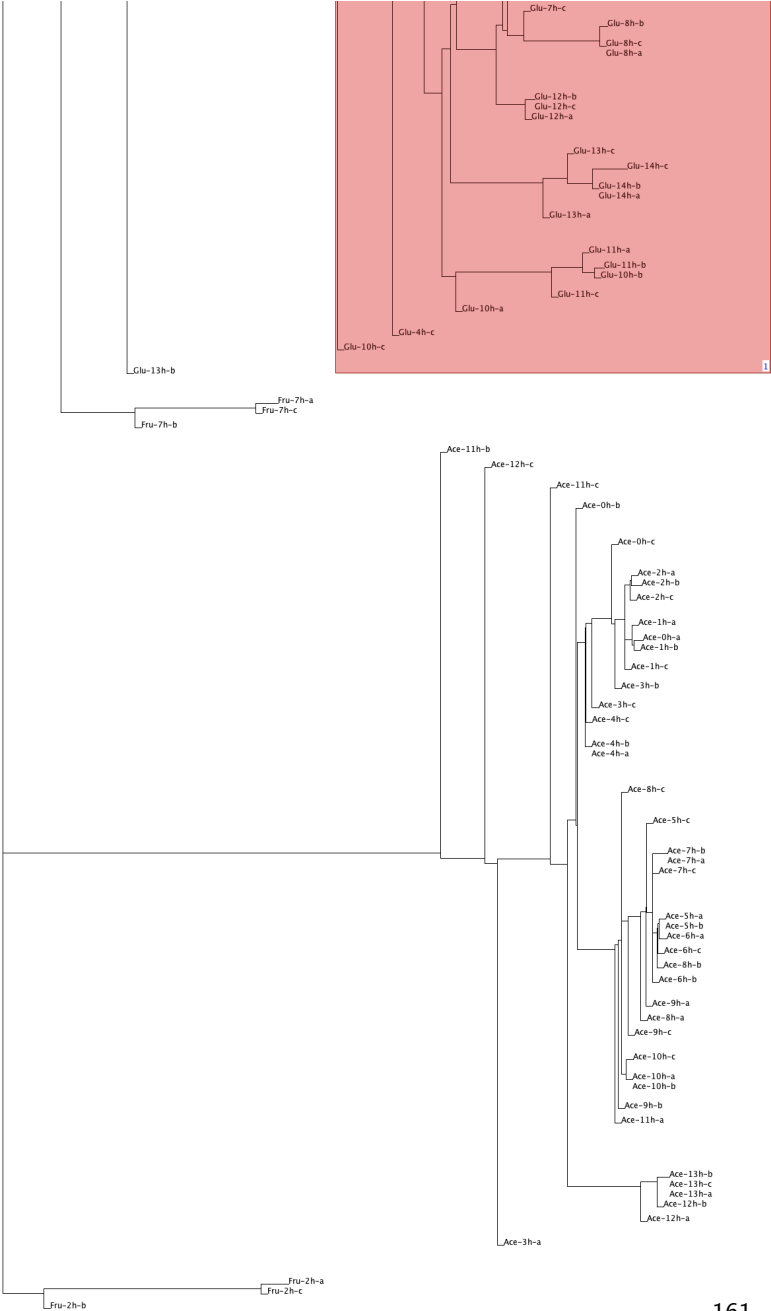
Zuletzt möchte ich mich bei Anne bedanken. Du warst fast seit

Beginn der Doktorarbeit an meiner Seite, hast mir oft eine Schulter zum anlehnen geboten, den Rücken gestärkt, Wege freigeräumt, Entscheidungen mitgetragen und mir mein Leben versüßt. Besonders dafür dass Du auch in den widrigen Phasen dieser Arbeit immer zu mir gehalten und mir Mut geschenkt hast gilt Dir mein aller tiefster und liebevollster Dank.

Anhang

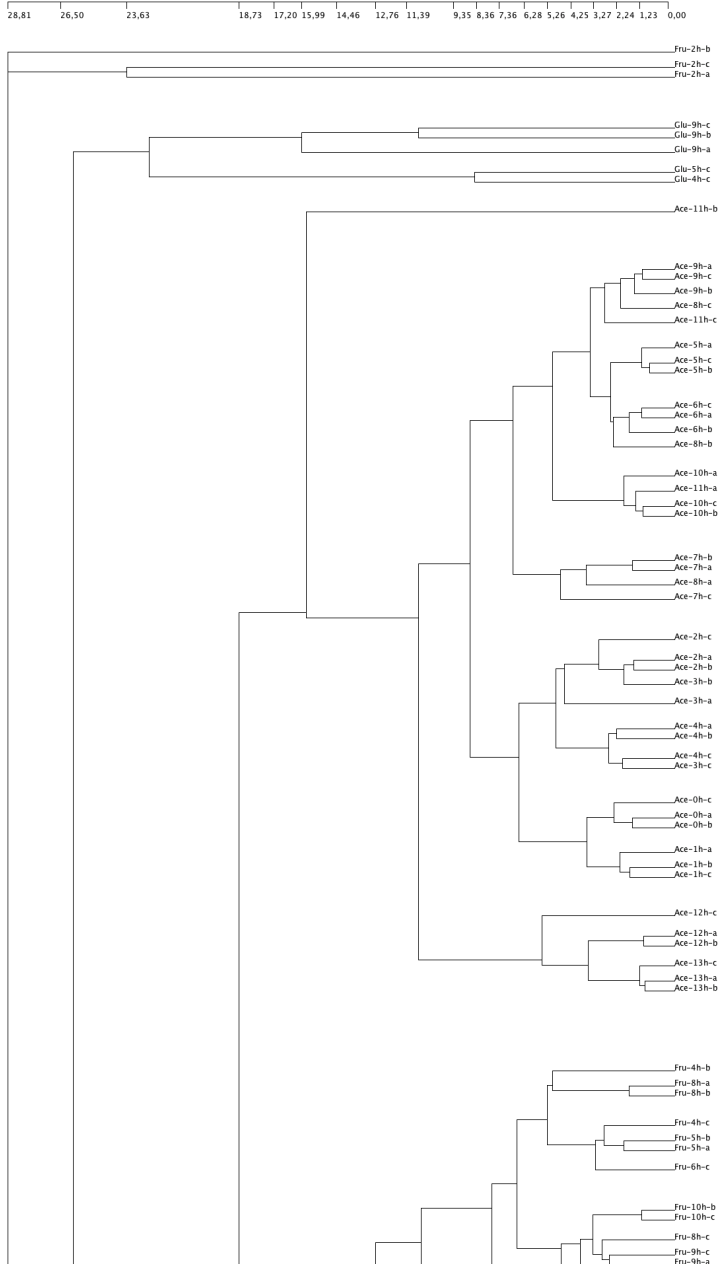
A1 Dendrogramm der ESOM

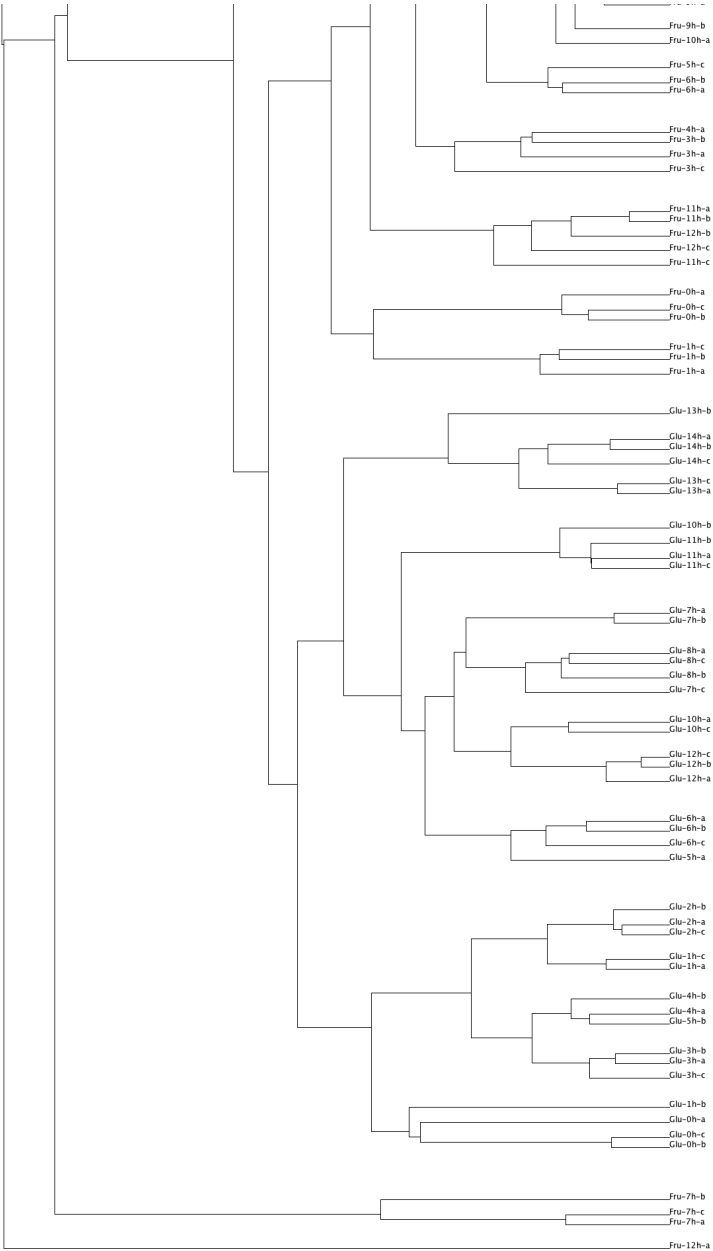




A2 Dendrogramm der HCA

A2 Dendrogramm der HCA





A3 Beweis zu Satz 1

Zu zeigen ist, dass

$$\frac{d\langle \mathbf{n}(t), \begin{pmatrix} 0 \\ 1 \\ \dots \\ s \end{pmatrix} \rangle}{dt} = 0$$

für Populationen, die sich gemäß

$$\frac{d\mathbf{n}(t)}{dt} = (-\mathbf{Q} + {}^+\mathbf{Q} - \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t)$$

verhalten, wahr ist.

Beweis: Es gilt,

$$\langle \mathbf{n}(t), \begin{pmatrix} 1 \\ 2 \\ \dots \\ s \end{pmatrix} \rangle = \sum_{i=1}^{s+1} \frac{d(n_i(t) \cdot (i-1))}{dt}$$

was sich nach ausklammern von $(i-1)$ umschreiben lässt als

$$\sum_{i=1}^{s+1} \frac{d(n_i(t) \cdot (i-1))}{dt} = \sum_{i=1}^{s+1} \left((i-1) \cdot \frac{dn_i(t)}{dt} \right).$$

Für $\frac{dn_i(t)}{dt}$ ergibt sich nach Ausmultiplizieren des zugehörigen Gleichungssystems

$$\frac{dn_i(t)}{dt} = \left[\sum_{j=1}^{s+1} ({}^+q_{i,j} + {}^-q_{i,j}) \cdot \mu_j \cdot n_j(t) \right] - \mu_i \cdot n_i(t)$$

und einsetzen in die vorletzte Gleichung:

$$\sum_{i=1}^{s+1} \left((i-1) \cdot \left[\sum_{j=1}^{s+1} [({}^+q_{i,j} + {}^-q_{i,j}) \cdot \mu_j \cdot n_j(t)] - \mu_i \cdot n_i(t) \right] \right) = 0.$$

Nach Addition der Gleichung mit $\sum_{i=1}^{s+1} (i-1) \cdot \mu_i \cdot n_i(t)$ ergibt sich

$$\begin{aligned}
 \sum_{i=1}^{s+1} (i-1) \cdot \mu_i \cdot n_i(t) &= \sum_{i=1}^{s+1} (i-1) \cdot \sum_{j=1}^{s+1} [(+q_{i,j} + -q_{i,j}) \cdot \mu_j \cdot n_j(t)] \\
 &= \sum_{i=1}^{s+1} \sum_{j=1}^{s+1} [(i-1) \cdot (+q_{i,j} + -q_{i,j}) \cdot \mu_j \cdot n_j(t)] \\
 &= \sum_{j=1}^{s+1} \sum_{i=1}^{s+1} [(i-1) \cdot (+q_{i,j} + -q_{i,j}) \cdot \mu_j \cdot n_j(t)] \\
 &= \sum_{j=1}^{s+1} \left[\sum_{i=1}^{s+1} (i-1) \cdot (+q_{i,j} + -q_{i,j}) \right] \cdot \mu_j \cdot n_j(t).
 \end{aligned}$$

Da i und j über identische Sequenzen $\{1, 2, \dots, s+1\}$ laufen, ist es zulässig diese auf der rechten Seite der Gleichung zu vertauschen:

$$\sum_{i=1}^{s+1} (i-1) \cdot \mu_i \cdot n_i(t) = \sum_{i=1}^{s+1} \left[\sum_{j=1}^{s+1} (j-1) \cdot (+q_{j,i} + -q_{j,i}) \right] \cdot \mu_i \cdot n_i(t)$$

Für den Rest des Beweises ist es ausreichend zu zeigen, dass die Gleichung

$$\sum_{j=1}^{s+1} (j-1) \cdot (+q_{j,i} + -q_{j,i}) = i-1 \quad \forall \quad i \in [1, s+1]$$

richtig ist. Aufgrund Eigenschaft 3.6 sind $+q_{j,i}$ und $-q_{j,i}$ für alle $j > i$ gleich 0 und können daher vernachlässigt werden:

$$\sum_{j=1}^i (j-1) \cdot (+q_{j,i} + -q_{j,i}) = i-1.$$

Desweiteren ist aufgrund Eigenschaft 3.7

$$\sum_{j=1}^i (j-1) \cdot {}^+q_{j,i} + (j-1) \cdot {}^+q_{i-j+1,i} = i-1$$

Der Koeffizient ${}^+q_{i-j+1,i}$ läuft über die Sequenz $j = (1, 2, \dots, i)$ rückwärts über ${}^+q_{j,i}$ von ${}^+q_{i,i}$ hinzu ${}^+q_{1,i}$. Daher ist die Permutation $(i-j+1) \rightarrow (j) \rightarrow (i-j+1)$ im rechten Teil der Summe zulässig, so dass:

$$\begin{aligned} i-1 &= \sum_{j=1}^i (j-1) \cdot {}^+q_{j,i} + ((i-j+1)-1) \cdot {}^+q_{j,i} \\ &= \sum_{j=1}^i (j-1 + (i-j+1) - 1) \cdot {}^+q_{j,i} \\ &= \sum_{j=1}^i (j-1) \cdot {}^+q_{j,i} \\ &= (i-1) \cdot \sum_{j=1}^i {}^+q_{j,i}. \end{aligned}$$

Wegen den Eigenschaften 3.5 und 3.6 gilt $\sum_{j=1}^i {}^+q_{j,i} = \sum_{j=1}^{s+1} {}^+q_{j,i} = 1$. ■

A3 Beweis zu Satz 2

Zu zeigen ist, dass

$$\frac{d \sum_i^{s+1} n_i(t)}{dt} = 0$$

für Populationen, die sich gemäß

$$\frac{d\mathbf{n}(t)}{dt} = (\mathbf{P} - \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t).$$

verhalten, wahr ist.

Beweis: Es gilt

$$\frac{d \sum_i^{s+1} n_i(t)}{dt} = \sum_i^{s+1} \frac{dn_i(t)}{dt} = \langle (\mathbf{P} - \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t), \mathbf{1} \rangle$$

Weiterhin ist:

$$\mathbf{P} - \mathbf{I} = \begin{pmatrix} -p_{2,1} & 0 & \dots & 0 & 0 \\ p_{2,1} & -p_{3,2} & \dots & 0 & 0 \\ 0 & p_{3,2} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & p_{s,s-1} & -p_{s+1,s} & 0 \\ 0 & 0 & \dots & p_{s+1,s} & 0 \end{pmatrix}.$$

Damit ist

$$(\mathbf{P} - \mathbf{I}) \boldsymbol{\mu}^D \mathbf{n}(t) = \begin{pmatrix} -p_{2,1} \cdot \mu_1 \cdot n_1 \\ p_{2,1} \cdot \mu_1 \cdot n_1 - p_{3,2} \cdot \mu_2 \cdot n_2 \\ \dots \\ p_{s,s-1} \cdot \mu_{s-1} \cdot n_{s-1} - p_{s+1,s} \cdot \mu_s \cdot n_s \\ p_{s+1,s} \cdot \mu_s \cdot n_s \end{pmatrix}.$$

Die Spaltensumme über diesem Ausdruck (und damit $\sum_i^{s+1} \frac{dn_i(t)}{dt}$) ist offensichtlich immer 0.

■

